

THIS WEEK



EDITORIALS

WORLD VIEW The woes facing space science and astronomy in the United States **p.127**

NOT SO EASY Waters of Mississippi fail to revive flagging coastal wetlands **p.128**

TREE TOXIN Bark gives African rat's predators a worse bite **p.128**

Who watches the watchmen?

Some commercial firms that oversee the ethics and scrutiny of clinical trials have been found wanting. Human volunteers in research deserve better.

Late last month, the US government proposed an overhaul of regulations covering research on human volunteers. It includes a move to allow single institutional review boards (IRBs) to oversee US multi-site clinical trials. Such boards underpin the ethics and safety of a trial, reviewing all research protocols and consent forms before it is launched. Currently, dozens or scores of IRBs — one for each trial site — can separately review and approve a clinical trial and its necessary paperwork, leading to massive duplication of effort and delays.

Commercial IRBs have multiplied in recent decades, reviewing and passing proposals for profit. Their creation was spurred partly by an explosion of industry-sponsored clinical trials, but they are increasingly passed business from academic centres, to relieve overstretched in-house ethics committees. There are now dozens of these companies and they will see potential business in the proposed shift to single-IRB oversight of multi-site trials. Commercial boards advertise themselves as faster at approvals than their academic counterparts. This promised speed and convenience could see organizers of multi-site trials turn to a commercial board rather than, for instance, the IRB at the principal investigator's institution.

Why does this matter? Commercial IRBs, like their academic counterparts, are as expert and committed as the people who sit on them. But the for-profit groups have an unsettling incentive to approve trials. If they are too tough, applicants may take their next proposal and their fees elsewhere. (In the proposed changes to the regulations, the government acknowledges that such 'IRB shopping' is a risk and asks for public comment on how it could be prevented.)

There is evidence that not all the commercial IRBs are as thorough as researchers, and their human subjects, would like. In the same week

that the government published the proposed rule changes, the Essex Institutional Review Board in Lebanon, New Jersey, was issued with a warning letter from the Food and Drug Administration (FDA).

A common way to test IRBs is to submit fictitious applications containing problems that should raise alarm bells. According to the FDA, when Essex received one of these sting applications, reportedly from journalists, it approved the trial without putting appropriate warnings — including of a serious heart risk and of a common risk of graft rejection — in the informed consent, or even discussing the heart risk, according to minutes of its meeting. Nor did it spot that both the trial sponsor and the lead investigator were invented. The company, which has done work for the National Institutes of Health's AIDS Clinical Trials Group, among others, received similar FDA warnings more than a decade ago.

The FDA has also censured other IRBs. Two years ago, the Coast IRB in Colorado Springs, Colorado, disbanded after the Government Accountability Office (GAO) submitted a phony clinical-trial application that the IRB approved. Some commercial IRBs carry out first-rate, conscientious work. Members of two other for-profit IRBs dismissed the phony GAO application as "awful" and "junk". One reviewer called it the "riskiest thing I've ever seen on this board".

And IRBs at academic institutions have their own ulterior motives to approve trials — among them institutional profile and prestige. And they, too, make mistakes. But the dollar remains the most powerful incentive to cut corners. The volunteers on whom clinical trials depend deserve the highest possible protection. Before implementing change that could see private companies overseeing an increased number of multi-site trials, the government should ensure that scrutiny of these IRBs is expanded to match. ■

NEON and on

The launch of an ecological monitoring network is good news at a difficult time.

After years of wrangling, scientists in the United States are set to wire up the natural landscape to a state-of-the-art environmental monitoring system. The National Ecological Observatory Network (NEON) will take Mother Nature's pulse and track her vital signs during both bad times and good.

In the coming years, NEON will pour massive quantities of biological and physical data from plains, forests and lakes into computer spreadsheets to look for trends (see page 135). It is a bold initiative whose time has clearly come, but it will take more than the flip of the on switch to make NEON glow.

The project's elegance is appealing, but behind its beguiling simplicity lurk a number of potential pitfalls. Some are technical: it would be foolhardy to underestimate the challenge of collecting hundreds of data streams in a uniform manner across dozens of sites and over the course of decades. Others are social and institutional: ecologists have long gathered data independently; NEON represents a dramatic shift towards the kind of collective data-driven system that physicists and astronomers have been pursuing for decades. How will it be received on the ground? And what's more, how many scientists will be ready to use the data when they come in?

The plan for NEON is to develop a healthy community of scientists who can make good use of the data. There are no guarantees, but the project is a risk worth taking. Human pressure on the biosphere will only escalate, and the sooner scientists get a firm handle on what that means the better. NEON could be a powerful tool for those investigations. And, in an exceedingly difficult fiscal environment, the new ecological network represents a refreshingly forward-looking initiative. Ecologists (and headline writers) everywhere should welcome it. ■



The dark clouds over US astronomy

The proposed cancellation of NASA's latest space telescope shows the difficulties ahead. But there is a solution, says Michael S. Turner.

It is barely 12 months since US astronomy was shown the future, with the release of *New Worlds, New Horizons in Astronomy and Astrophysics*, the latest decadal survey by the National Academy of Sciences. The survey offered a strategy — based on realistic budgets and leveraged by international and private partnerships — to realize dazzling opportunities, including searching for life on other planets, identifying dark matter and understanding dark energy. It also promised to reveal the evolution of the first stars and galaxies and to probe whether supermassive black holes are accurately described by general relativity.

It began well, with the Large Synoptic Survey Telescope, the highest priority for ground-based astronomy, put on a path for construction in 2014. But today, US astronomy is in disarray, community morale is low and the proposed cancellation of the James Webb Space Telescope (JWST), the cornerstone of the 2010 decadal survey, is the most glaring example of difficulties ahead. Although capabilities and budgets in Europe and Asia have largely caught up with those of the United States, the health of the US programme remains critical, not just domestically, but also to the rest of the world.

As the opportunities have soared and projects have become more complex and costly, government spending on science has been shrinking. NASA's science budget is in steady decline. The House of Representatives has said that it wants to cut another 10% from NASA for 2012 — and that was before the bill to cut domestic spending over the next decade was passed earlier this month.

This has undermined NASA's planned collaboration with the European Space Agency (ESA) on the International X-ray Observatory (IXO) and on a space-based interferometer to detect gravity waves — the Laser Interferometer Space Antenna (LISA). Both partnerships were endorsed by the decadal survey, but the Europeans are now looking at cheaper and less-ambitious missions that they can do alone.

The decadal survey's top priority was the Wide Field Infrared Survey Telescope (WFIRST), designed to find exoplanets and probe dark energy, but budget trouble has pushed its earliest possible launch into the mid-2020s. By then, much of its key science could have been done by a similar ESA mission, Euclid. (Euclid is competing for a slot in ESA's Cosmic Vision programme, with an announcement expected this October.)

The JWST illustrates the complexity of the problems in the US astrophysics programme. With a 6.5-metre deployable mirror and powerful infrared instruments, it will be a hundred times more sensitive than the Hubble Space Telescope and a worthy successor. It will image the first stars and galaxies, and search for water and evidence of life on distant planets. Yet it is

at least US\$1.6 billion over budget, and last month the House voted to terminate it.

The JWST can still be saved, and should be. The budget for 2012 is yet to be settled and there is a groundswell of support for the telescope: professional societies around the globe have issued statements of support, online petitions have been organized and influential commentators have called for its completion. But if the JWST does launch, its cost could yet wreak havoc on the rest of the astrophysics programme.

I believe that there is a path forward for astronomy. It will involve hard choices and better international coordination, but the scientific payoffs ahead demand that we try.

First, it is essential that the United States finishes the JWST. As well as being the cornerstone of the future astrophysics programme, the

telescope will transform global astronomy just as Hubble has. But NASA must spread the additional cost across the agency so that it does not cripple the rest of US astrophysics. This will not be easy, but without a clear plan for the post-shuttle era, NASA is going to need a steady stream of exciting science results to capture the public's imagination and loosen the purse strings of Congress.

Second, ESA and NASA will have another opportunity to merge their competing dark-energy missions, Euclid and WFIRST, if Euclid is selected for a Cosmic Vision slot. The agencies should do so, and next year they should seek a renewed partnership on IXO and LISA.

Third, agencies on both sides of the Atlantic have their sights set on a giant ground-based telescope, 30 metres or more in aperture, with complex adaptive optics to complement the

JWST. Both sides are seeking international and private partners. A coherent strategy to build at least two of these grand machines of discovery — one for the northern sky and one for the southern — is essential. In the United States, the National Science Foundation must clarify the currently muddled situation by picking which of the competing private projects it wants for a partner. And scientists around the world must work together to develop the necessary technology and instrumentation for these billion-dollar telescopes.

We are in an unprecedented age of discovery, with opportunities to answer the biggest questions humankind can ask. Where did we come from? Where are we going? Are we alone? If the global science community can figure out how to live on leaner budgets by working together more effectively, this could yet be a golden age for astronomy. ■

Michael S. Turner is Rauner distinguished service professor and director of the Kavli Institute for Cosmological Physics at the University of Chicago, Illinois.
e-mail: mturner@uchicago.edu

**NASA IS GOING TO
NEED A STEADY
STREAM OF
EXCITING
SCIENCE RESULTS
TO CAPTURE
THE PUBLIC'S
IMAGINATION.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/tbcucn

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

STATISTICS

Ten thousand Kims of Korea

Kim is the most common surname in Korea, and has probably held pole position for more than a millennium, according to a statistical study.

Petter Minnhagen at Umeå University in Sweden and his colleagues studied ten Korean family trees that traced lineages back through 500 years.

Using a statistical analysis that divides the tree data into family groups and predicts how they mix, the team was able to extrapolate the frequency of the name Kim back to AD 500. The model predicts that at that time 10,000 inhabitants — about one-fifth of the total population — shared the surname.

The results suggest that Korean culture has been stable for hundreds of years, the authors say.

N. J. Phys. 13, 073036 (2011)

ZOOLOGY

Poisonous surprise for predators

By chewing on the bark of a poisonous tree, the African crested rat acquires a toxin that it delivers if bitten by predators — the first placental mammal found to combat predation in this way.

When threatened, the rat (*Lophiomys imhausi*) parts its grey fur to expose a pattern of

specialized black and white hairs on its flanks (**pictured**), daring its attacker to bite the target. The creature picks up the toxin with its saliva by gnawing on the *Acokanthera schimperi* tree, then drools on the hairs.

Fritz Vollrath at the University of Oxford, UK, and his colleagues examined the structure of these hairs under two types of electron microscope. They found the hairs to be porous cylinders

enclosing many long fibrils — a structure well adapted to absorbing the toxin and to delivering it upon contact with a predator's mouth.

Proc. R. Soc. B <http://dx.doi.org/10.1098/rspb.2011.1169> (2011)

METABOLISM

Hungry cells eat from within

When food is scarce, some brain cells begin to devour themselves, activating an appetite-stimulating

molecule in the process.

A brain region called the hypothalamus contains neurons that regulate feeding in response to nutritional signals. Rajat Singh and his colleagues at the Albert Einstein College of Medicine in the Bronx, New York, found that these neurons cannibalize their own organelles, proteins and lipid stores — a process known as autophagy — when starved of nutrients. This liberates fatty acids that cause the cells to express higher levels of an appetite-stimulating



J. DERMANSKY/CORBIS

CONSERVATION

Louisiana marsh restoration has failed

For two decades, fresh water from the Mississippi River has been diverted into Louisiana's shrinking coastal wetlands (pictured) in an attempt to restore them, but three such projects have failed to increase vegetation or marsh area. Moreover, the diversions have raised the marshes' risk of suffering hurricane damage.

Michael Kearney at the University of Maryland in College Park and his team analysed satellite images collected between 1984 and 2009 for three of the longest-running diversion projects. They found that most new plant growth

consisted of floating plants and algae rather than deep-rooted marsh plants that hold soil in place. For one of the diversions, Hurricane Katrina destroyed the most vegetation in zones that received the most direct input of fresh water.

The authors say that an influx of nutrients in the river water, probably from agricultural run-off and industry, has been damaging marsh plants' roots.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2011gl047847> (2011)

For a longer story on this research, see go.nature.com/qticwp



R. SOC.

compound called agouti-related peptide.

When the researchers deleted a gene required for autophagy in the mouse neurons that produce this peptide, the mice ate less and were slimmer than control animals with normal autophagy.

Cell Metab. 14, 173–183 (2011)

MICROBIOLOGY

Biofilm for yeast sex

Some microorganisms stick together to form impenetrable, drug-resistant biofilms that can contaminate surfaces such as those of medical devices. The yeast *Candida albicans* is one such microbe, but certain strains form another type of biofilm — one that is susceptible to drugs and is regulated by a different molecular pathway.

David Soll and his colleagues at the University of Iowa in Iowa City found that about 10% of *C. albicans* strains form this second type of biofilm, which is also permeable to immune cells. These strains have a genetic difference at one locus that allows the yeast to reproduce sexually. The authors suggest that this minority type of biofilm aids in the mating process by mediating the movement of hormone signals.

Comparing the different pathways that regulate the two types of biofilm could yield clues about what makes a biofilm drug-susceptible, which could aid in the development of treatments to disrupt harmful biofilms.

PLoS Biol. 9, e1001117 (2011)

CELL BIOLOGY

Alzheimer's in a dish

Skin cells taken from patients with Alzheimer's disease and converted into neurons exhibit molecular defects that have been linked to the neurodegenerative condition.

Such cells could function as models for those working to understand the disease.

Asa Abeliovich at Columbia University in New York and his colleagues created neurons from the fibroblasts of three patients with genetic mutations that cause a rare hereditary form of Alzheimer's disease. The authors used retroviruses to introduce key genes into the skin cells, directly reprogramming them into neurons.

Compared with converted cells from healthy people, those from patients with disease-related mutations produced higher levels of a protein fragment called amyloid- β , which is linked to Alzheimer's disease. All reprogrammed neurons contained amyloid precursor protein — which is clipped to make amyloid- β — but in cells from people with Alzheimer's mutations this protein was more localized in membrane-bound sacs.

Cell 146, 359–371 (2011)

MATERIALS

Fabric that puts out fire

A specially treated cotton fabric can actively extinguish flames when it is set on fire.

The material was produced by Jaime Grunlan at Texas A&M University in College Station and his colleagues by dipping the fabric repeatedly into two different solutions of water-soluble polymers, creating a coating just hundreds of nanometres thick. When heated to high temperatures, this layered, 'intumescent' film reacts to form a carbon-based foam that lowers the amount of gas fuelling the burn and slows heat transfer, inhibiting ignition. Flame testing showed that whereas the material burned when coated with just five protective layers (pictured, left panel), 20 layers were sufficient for it to put out the flame soon after ignition (right panel).

Such coatings may provide

COMMUNITY CHOICE

The most viewed papers in science

CLIMATE

Hot cities heat up China

HIGHLY READ
on www.agu.org
the week of 1 Aug

Urbanization is responsible for roughly 24% of the warming observed in eastern China during the past three decades, thanks to the build up of heat in cities — known as the urban heat island effect.

Baode Chen at the Shanghai Typhoon Institute and his colleagues analysed monthly mean temperature data from 463 meteorological stations in a range of urban and rural settings between 1981 and 2007. Using both raw and reanalysed data, the team found that urbanization contributed up to 44% of the temperature increase (almost 0.4 °C per decade) in metropolises with populations of more than 1 million. The largest increase in the urban heat island effect came after the early 2000s, coinciding with a period of rapid growth of cities in China.

J. Geophys. Res. <http://dx.doi.org/10.1029/2010JD015452> (2011)



an alternative to chlorinated or brominated flame-retardant chemicals, which persist in the environment if they leach from fabrics, and show signs of toxicity in animal studies.

Adv. Mater. <http://dx.doi.org/10.1002/adma.201101871> (2011)

DEVELOPMENT

Placenta protects the brain

Far from being simply a passive supplier of food, the mouse placenta protects the developing fetus from brain damage if the mother is deprived of food.

Kevin Broad and Barry Keverne at the University

of Cambridge, UK, studied developing embryos and placentas taken from mothers that had gone 24 hours without food about mid-way through gestation — a stage in pregnancy that is crucial for the development of a brain region called the hypothalamus. The authors found shifts in gene expression, including upregulation in the placenta of genes involved in autophagy, a process by which cells degrade their own components. Meanwhile, in the hypothalamus the expression of genes linked to neural development was enhanced.

The authors suggest that the placenta breaks down its own tissues, recycling proteins inside its cells to provide a steady supply of nutrients to the developing hypothalamus.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1106022108> (2011)

For a longer story on this research, see go.nature.com/ej2kxh

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Science's pay gap

In the United States, the pay gap between men and women is slightly smaller in science, technology, engineering and mathematics (STEM) than in other fields, according to a report released on 3 August by the US Department of Commerce. Women's average hourly earnings are 14% lower than men's in STEM jobs, but 21% lower in non-STEM jobs. When the data are adjusted to control for other factors, such as education, the STEM pay gap shrinks to 12%, with physical and life sciences coming in at 8%. Women comprise 24% of STEM workers, a proportion that has remained constant over a decade. See go.nature.com/mshimp for more.

To Jupiter!

NASA's Juno spacecraft launched successfully on 5 August, bound for Jupiter. The US\$1.1-billion mission will take five years to reach the Solar System's largest planet, which it will swing round 33 times in a highly elliptical orbit. Juno will probe the Jovian atmosphere for water vapour, and look for signs of a solid core (see *Nature* **476**, 13–14; 2011).

China's sea claims

Tensions increased between Vietnam and China over exploration in the South China Sea, after China's state media reported on 2 August that the research vessel *Tan Bao Hao* had surveyed the sea's southwest basin in late July, with the aim of studying tectonic activity. Vietnam's foreign ministry demanded that China "cease and refrain" from further such acts.

The two countries — and other neighbouring nations such as the Philippines — are

in dispute over their claims to potentially oil-rich islands in the region. China's recent funding boost for deep-sea exploration is largely motivated by the nation's desire to find oil and minerals (see *Nature* **466**, 166; 2010).

NSF gets social

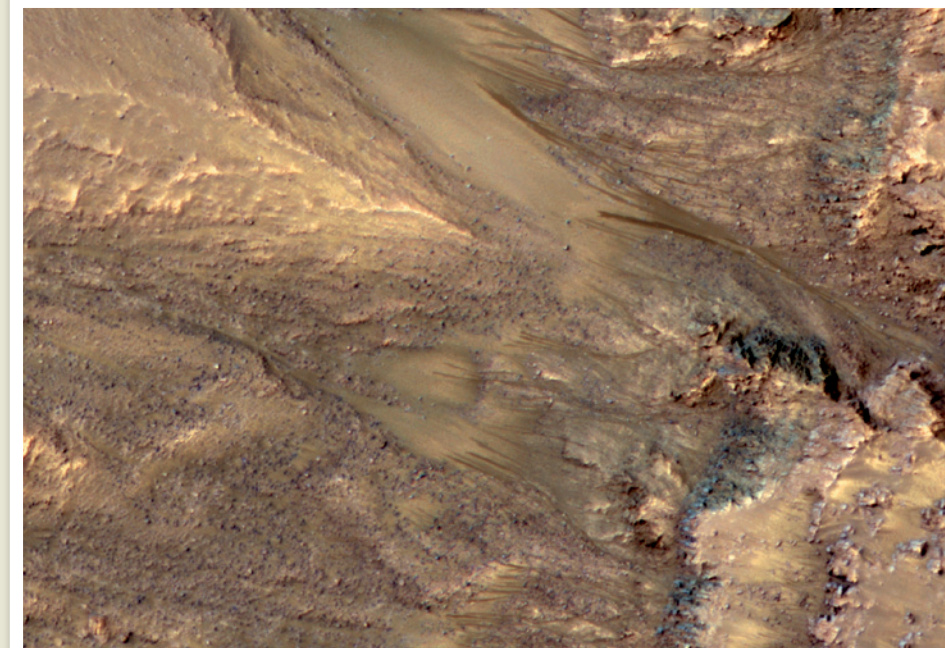
The fourth of the cross-disciplinary 'synthesis' centres set up by the US National Science Foundation (NSF) will be the first to involve social scientists. So far, synthesis centres founded in 1995, 2004 and 2008 have focused, respectively, on ecology, evolutionary biology and the interface

between mathematics and biology. The National Socio-Environmental Synthesis Center, in Annapolis, Maryland, will research issues such as food production, water availability and the interaction between human activities and ecosystem health, the NSF says. It announced funding of US\$27.5 million over 5 years for the centre on 2 August.

Ecology network

Researchers behind a long-planned US\$434-million environmental monitoring network released the project's draft scientific plan on 4 August. The National Ecological Observatory

thousands of narrow, dark streaks that appear on some of Mars's steep slopes during warmer seasons (pictured, in Newton Crater). By winter, the streaks have faded or vanished. The findings were published on 5 August (A. S. McEwen *et al. Science* **333**, 740–743; 2011). See go.nature.com/casgrb for more.



NASA/JPL-CALTECH/UNIV. ARIZONA

Streaks suggest liquid water on Mars

NASA scientists have reported evidence of salty water near the surface of Mars, in a region that may become a landing target for future missions to determine whether the planet can support life. Images from the High Resolution Imaging Science Experiment (HiRISE), a camera aboard the Mars Reconnaissance Orbiter, revealed

Network, which aims to collect data for 20 'core' observatories representing ecosystems across the United States, had cleared its final funding hurdle a week earlier, when the National Science Foundation awarded money to begin construction. See page 135 for more.

PEOPLE

Stem-cell chief

The US National Institutes of Health has finally found a director for its Intramural Center for Regenerative Medicine, a US\$52-million, 7-year initiative begun in 2010. Mahendra Rao, a

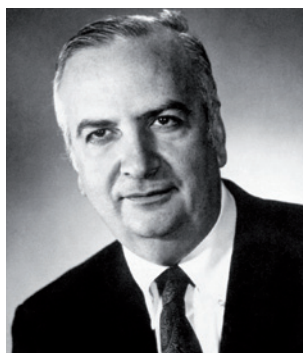
A stem-cell biologist who spent six years as vice-president of regenerative medicine at Life Technologies (formerly Invitrogen) in Carlsbad, California, will take up the position later this month. His appointment was announced on 3 August. See go.nature.com/urclckh for more.

Ex-NIH head dies

Bernadine Healy, who was the first woman to lead the US National Institutes of Health (NIH), died on 6 August aged 67. Healy, who trained as a cardiologist, headed up the NIH between 1991 and 1993, taking over after the agency had been without a director for two years. Her greatest impact was in championing research into women's health — she launched the Women's Health Initiative, a US\$625-million clinical trial which, in 2002, showed that hormone-replacement therapy increases the risk of breast cancer, stroke and heart disease.

Nobel laureate dies

Immunologist Baruj Benacerraf (pictured), who shared the 1980 Nobel Prize in Physiology or Medicine for his work on the genetic basis of the immune system, died on 2 August, aged 90. Benacerraf was born in Venezuela and brought up in France, but it was at New York University in the 1950s and 1960s that



he traced how guinea pigs' varying immune responses to the same antigen depended on specific genes. From 1980 to 1992, he was president of the Dana-Farber Cancer Institute in Boston, Massachusetts.

POLICY

Fukushima's wake

Japan will next year create a more independent nuclear-safety agency, scrapping its existing watchdogs — the Nuclear and Industrial Safety Agency (NISA) and the Nuclear Safety Commission. Both currently sit within the pro-nuclear trade and industry ministry, but the new watchdog would sever ties with that ministry and might impose tougher regulations on nuclear plants, said Goshi Hosono, the minister in charge of the country's response to the Fukushima meltdowns. He announced the plan on 5 August; one

day after the government had fired NISA's director and two other industry ministry officials.

Nuclear closure

Only two commercial facilities in the world recycle plutonium and uranium from used nuclear fuel to make 'mixed oxide' (MOX) fuel: now one of them is shutting down. Britain's Nuclear Decommissioning Authority on 3 August announced the closure of the MOX facility at Sellafield. The plant was supposed to supply Japanese reactors with MOX, but has been left without a market in the aftermath of the nuclear meltdown at the Fukushima Daiichi plant. See page 140 for more.

Integrity plans

The US Environmental Protection Agency's draft policy on scientific integrity, released on 5 August, drew immediate criticism. The watchdog group Public Employees for Environmental Responsibility, headquartered in Washington DC, called it "by far the weakest scientific integrity rules of any agency". Among its concerns are that managers and public-affairs staff can screen and approve the release of scientific information. The US National Science Foundation released a draft policy on 4 August. All

COMING UP

11–12 AUGUST

A US Institute of Medicine committee examining the need for chimpanzee research holds a public workshop in Washington DC. The group is expected to release a report by the end of the year.

go.nature.com/obxi2t

11–18 AUGUST

Results from detectors looking at cosmic rays — including the International Space Station's Alpha Magnetic Spectrometer — should be revealed at the 32nd International Cosmic Ray Conference in Beijing.

<http://icrc2011.ihep.ac.cn>

12–14 AUGUST

Expect updates on the attempt to build an elevator into space, from the 2011 Space Elevator Conference in Redmond, Washington.

go.nature.com/krqd8l

agencies had to submit policies to the White House Office of Science and Technology Policy by 5 August. See go.nature.com/m2q2zg for more.

Data mining OK

An update to Britain's copyright laws means that scientists will be allowed — for research purposes — to mine text and data from journal articles, databases and other literature without having to ask permission from the copyright owner. Such text digitization and data mining gets a specific exemption under government plans to modernize the country's intellectual-property laws, announced on 3 August. See go.nature.com/hekfgw for more.

► NATURE.COM

For daily news updates see:

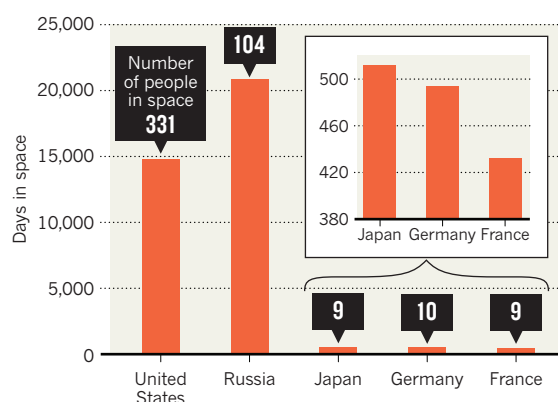
www.nature.com/news

TREND WATCH

Japanese astronauts clocked their collective 494th day in space on 24 July, passing Germans, with 493, to make Japan the third most space-faring country. As of 10 August it had reached 511 days. Japan has no hope of catching up with Russia's 20,817 days or the United States' 14,820, but International Space Station schedules show that Japanese astronauts will keep a firm grip on third place. China will become the next contender for high honours later this decade, as it establishes its own space station.

A NEW ORDER IN SPACE

Japanese astronauts have passed 500 days in space — putting the nation third in the space-faring rankings.



NEWS IN FOCUS

ANCIENT DNA Pinpointing the legacy of trysts with extinct relatives **p.136**

CHINA Prestigious academies vow to clean up process for electing members **p.139**

NUCLEAR POWER One option gone for UK plutonium stockpile **p.140**



MEDICINE Debating the need for sham surgery in Parkinson's trials **p.142**

M. WILSON/GETTY



Political pressure to balance the US federal budget will have an impact on research.

US BUDGET

Debt deal sets day of reckoning

US science agencies avoid immediate pain but could be devastated by automatic cuts in 2013.

BY ERIC HAND

Scalpel or guillotine? Those are the possible fates in store for US science funding after Congress and the White House reached a deal to cut federal spending and raise the nation's self-imposed debt limit before a 2 August deadline.

The product of tumultuous negotiations, the deal largely spares science in the short term but puts a day of reckoning on the horizon: 2 January 2013. If politicians cannot agree on

how to improve the government's fiscal outlook by then through targeted cuts and other means — the scalpel option — their failure will automatically trigger the guillotine: a deep cut applied across a range of expenditures, including research. In the worst case, the automatic cuts could mean shuttered laboratories and mass lay-offs at universities.

Whether or not that happens depends partly on a special Joint Select Committee on Deficit Reduction, the 'super-committee', to be set up as part of the deal. The super-committee,

which congressional leaders must appoint by 16 August, will be charged with finding ways of raising revenues or reducing the costs of entitlement programmes such as health care and social security, which the government is legally required to fund. If the super-committee fails, the automatic triggers would force discretionary spending to fall under the indiscriminate blade of across-the-board cuts.

"Then there will be extraordinary pain," says Michael Lubell, director of public affairs for the American Physical Society in Washington DC. "And it will get worse in 2014."

The two-stage structure of the debt deal explains both the short-term reprieve and the long-term worry. The first set of agreed cuts, totalling US\$917 billion, will be spread over 10 years, but two factors mitigate their effect. First, reductions to defence spending will account for a significant share of the cuts — meaning that other US agencies won't bear the entire burden. Second, the cuts are heavily loaded forward onto the 2014 fiscal year and beyond, in an apparent effort to shelter the current fragile economy. Only minimal cuts will be implemented in fiscal years 2012 and 2013.

In fact, the cap for overall discretionary spending that legislators will use in finishing the 2012 appropriations process is \$29 billion higher than the cap that House Republicans have used for their budget proposals so far (see 'Fiscal tug-of-war'). This means that the National Institutes of Health (NIH) might yet escape cuts to its funding that previously seemed likely. The higher cap also leaves wiggle room for Senate legislators to restore pet projects, such as NASA's James Webb Space Telescope, which the House last month voted to cancel.

The second set of cuts, totalling at least \$1.2 trillion, could inflict much deeper wounds. The task of selecting precisely how to make up that amount will fall to the super-committee, a group of six Democrats and six Republicans, who will have the power to look beyond discretionary funding to consider entitlement cuts and new revenue streams from taxes. The super-committee has until 23 November this year to formulate a plan, which will be put before Congress a month later. If the super-committee can't reach agreement, or if Congress won't approve its plan, the ►

► **NATURE.COM**

For ongoing US budget coverage see: go.nature.com/eg2noo

FISCAL TUG-OF-WAR

Significant gaps between President Obama's 2012 budget request and amounts approved by the House appropriators have yet to be reconciled, and some key questions answered, but the 2012 budget is far less painful for US science than what may lie ahead.

Agency	2011 enacted (US\$ million)	2012 president's request (US\$ million)	2012 House spending bill (US\$ million)	Outstanding issues
National Institutes of Health	30,766	31,829	Not yet put forward	Will the proposed National Center for Advancing Translational Sciences be approved by 1 October 2012?
NASA	18,448	18,724	16,810	Will the US Senate restore funding to the over-budget James Webb Space Telescope?
National Science Foundation	6,807	7,768	6,860	How low will grant-acceptance rates go? A short-term increase from 2009 stimulus funding is coming to an end.
Department of Energy Office of Science	4,858	5,416	4,800	Does the Advanced Research Projects Agency – Energy programme have a future? It barely survived in the House, after a floor amendment was accepted by one vote.
National Oceanic and Atmospheric Administration	4,597	5,498	4,485	Will the administration get a National Climate Service, a proposed consolidation of various climate-research activities?
National Institute of Standards and Technology	750	1,004	701	Will the Technology Innovation Program survive? The House would eliminate the \$75-million industry partner programme, which didn't offer new grants in 2011.

SOURCE: HOUSE APPROPRIATIONS COMMITTEE

COUNTDOWN TO CUTS

House speaker John Boehner and President Barack Obama face 18 months of budget wrangling during an election campaign.



16 AUGUST 2011

Congressional 'super-committee' selected.

30 SEPTEMBER 2011

End of 2011 fiscal year and current budget. Stop-gap legislation is likely until a 2012 budget is passed.

23 NOVEMBER 2011

Super-committee proposes how to reduce budget by US\$1.2 trillion.

23 DECEMBER 2011

Congress votes on super-committee's proposal.

FEBRUARY 2012

President Obama tables 2013 budget request.

6 NOVEMBER 2012

Presidential and congressional elections.

2 JANUARY 2013

If super-committee proposals not accepted, across-the-board cuts to all federal agencies begin.

► across-the-board \$1.2-trillion cut will automatically kick in a year later (see 'Countdown to cuts').

Such an outcome would be catastrophic, says Lubell. According to his back-of-the-envelope calculation, the automatic cut would slash funding for science agencies by 11%, starting in 2013. For the Department of Energy's \$5-billion Office of Science, even shutting down a national laboratory — for example Fermilab, the particle-physics laboratory in Batavia, Illinois, with its \$300-million annual budget — would achieve only part of the mandated savings. Granting agencies such as the NIH and the National Science Foundation (NSF) would have to lower their grant-acceptance rates to single digits. The knock-on effect would be catastrophic for universities, which depend on grants to help cover post doctoral researchers and infrastructure costs. "Can universities make up that gap? Not a chance," says Barry Toiv, spokesman for the Association of American Universities, based in Washington DC, an advocate for the largest research universities.

The global cut would also wreak havoc with the 2013 appropriations process, which would begin with President Barack Obama's budget request in February 2012. In such a scenario, appropriators could play favourites among different science agencies, boosting some at the expense of others in anticipation of the looming cuts, says Lubell.

Science agencies pursuing basic research that both parties tend to support, such as the NIH, the NSF and the core programme of the National Institute of Standards and Technology — could fare well. Basic research at the Department of Defense might also survive, as it offers little in the way of potential savings, being such a small piece of the overall defence pie.

But 'mission' agencies, such as NASA, or the National Oceanic and Atmospheric Administration (NOAA), where science is a large but ultimately ancillary activity, could suffer. Worst off would be programmes supporting research

that is controversial in the current Congress, says Patrick Clemins, director of the research and development budget programme at the American Association for the Advancement of Science in Washington DC. This could spell trouble for climate-change research at NOAA and within NASA's Earth Science Division. The Department of Energy's applied energy and loan-guarantee programmes could also be vulnerable, as some members of Congress see them as picking winners in a part of the economy that should be governed by the free market.

This is why science advocates are hoping that the make-up of the super-committee will favour the kind of negotiation that would head off across-the-board cuts. Toiv would like to see members selected from the Bowles-Simpson commission, which was appointed by Obama to study options for reducing the national debt. It emphasized infrastructure, education and basic research as important to the nation's long-term economic health. But Lubell expects the super-committee to be entrenched along partisan lines: entitlement reform would be off-limits for the Democrats, whereas new taxes would be anathema for Republican members. Refusing to compromise would enable both parties to pander to their traditional constituencies in the November 2012 general elections.

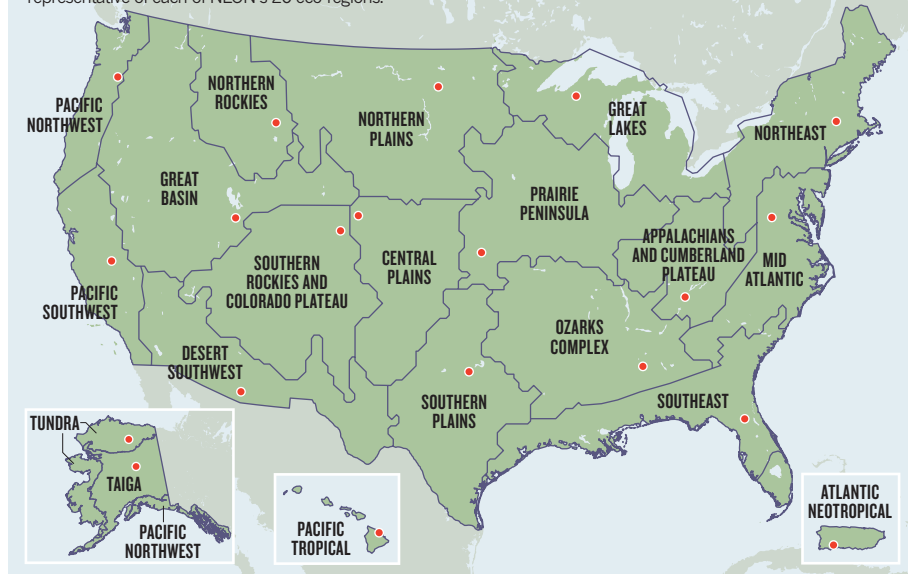
April Burke, president of Lewis-Burke Associates in Washington DC, a science-lobbying firm, is more sanguine, saying that an impasse would not necessarily trigger doomsday cuts. Congress, a slippery beast, has devised ways to avoid automatic triggers before. And the 2 January 2013 implementation of the mandated cuts is a political lifetime away. By then, the 2012 election could have changed the political stage, along with its players. The economy could improve — or get worse. And new rules could be written.

"I don't think we're looking at the full menu of what might happen then," she says. "I think there might be yet another reality." ■

AP/P. MARTINEZ MONSIVAS

GROUND COVER

Permanent research stations (red circles) are broadly representative of each of NEON's 20 eco-regions.



ECOLOGY

US launches eco-network

Ambitious project to systematically monitor the environment on a continental scale is finally ready to break ground.

BY JEFF TOLLEFSON

Ready or not, the era of big data is coming to ecology. After years of discussion and debate, the United States is moving forward with an environmental monitoring network that promises to help transform a traditionally small-scale, local science into a continental-scale group enterprise.

The National Ecological Observatory Network (NEON) will consist of 20 'core' observatories representing distinct eco-regions throughout the United States (see map). These will be bolstered by temporary stations that can be relocated wherever data need to be collected. The sites will house equipment and host visiting researchers, while gathering a range of environmental data over at least three decades.

The result will be a vast database that scientists can mine to tackle broad questions such as how global warming, pollution and land-use change are affecting ecosystems across the country. "NEON is really about trying to understand the biology of the entire continent rather than the biology of a specific place," says David Schimel, the project's chief

science officer, based in Boulder, Colorado.

Conceived more than a decade ago, NEON has already spent just over US\$80 million planning the network and developing instruments, and has a staff of about 140, including some 60 scientists and engineers. But the project, which is run by an independent body — NEON, Inc. — didn't clear its final hurdle until 28 July, when the National Science Foundation (NSF) awarded it \$434 million over the next decade, with \$12.5 million to be spent in the current fiscal year.

The money could jump-start site preparation and construction as early as this year. Schimel says that NEON expects to begin work near its home base in Colorado and in the northeast and to expand from there. The first data, from sensors placed on towers, submerged in streams and buried in soils, should come next year — a trickle that project members hope will become a torrent by 2016,

"Being able to manage and process and make sense of those data is going to be a huge challenge."

when the project will be fully operational.

"It's a huge step forward," says Sandy Andelman, an ecologist who heads a network of tropical ecology sites through Conservation International in Arlington, Virginia. But, she cautions, NEON will create a massive new data set that few environmental scientists know how to use today. "Being able to manage and process and make sense of those data is going to be a huge challenge," Andelman says.

Once the entire network is up and running, some 15,000 sensors will work in concert with scientists on the ground to supply roughly 500 distinct categories of data ranging from basic weather readings to concentrations of ozone in the air and nitrogen in the soils, leaves and streams. Scientists will collect tens of thousands of samples, including soil, water, plants and small mammals. At the same time, aerial surveys will analyse broader land-use trends as well as details such as leaf chemistry and carbon stocks, and satellite data will expand coverage over the entire continent.

NEON's draft scientific plan, released on 4 August, identifies climate, land use and invasive species as drivers of ecosystem change that can be studied through their impact on biodiversity, biogeochemical and hydrological cycles and the spread of infectious diseases. By gathering data over large territories and long periods of time, NEON will aim to give scientists the statistical power they need to tease out subtle trends in a shifting ecological balance. For this reason, rather than focusing on regional investigations, the research stations are mainly designed to gather data in a uniform manner that can be used in larger network studies.

"The big intellectual hurdle was the transition from a series of largely regional sites to an integrated system," says James Collins, an ecologist at Arizona State University in Tempe, who as head of the biological sciences directorate at the NSF during 2005–09 helped to mould the proposal into its current form.

Over the years the project has battled against scepticism. Early on, Steven Wofsy, an atmospheric chemist at Harvard University in Cambridge, Massachusetts, and a pioneer of carbon studies at the Harvard Forest ecological research centre, feared that NEON would generate more data than value. Today he gives NEON credit for putting high-profile critics — including himself — on its scientific advisory board to address any problems head on.

Wofsy remains sceptical of big science projects that promise to open up a new field with lots of data that scientists haven't necessarily said they want. But he commends NEON for developing a plan to tackle scientific questions that will play out over decades.

"If these guys are successful, they will have the goods on some really big ecological questions," Wofsy says. "Scientists who are interested in addressing those problems will have to make the effort to learn how to use those data." ■ SEE EDITORIAL P.125

GENOMICS

Ancient DNA reveals secrets of human history

Modern humans may have picked up key genes from extinct relatives.

BY EWEN CALLAWAY

For a field that relies on fossils that have lain undisturbed for tens of thousands of years, ancient human genomics is moving at breakneck speed. Barely a year after the publication of the genomes of Neanderthals¹ and of an extinct human population from Siberia², scientists are racing to apply the work to answer questions about human evolution and history that would have been unfathomable just a few years ago.

The past months have seen a swathe of discoveries, from details about when Neanderthals and humans interbred, to the important disease-fighting genes that humans now have as a result of those trysts.

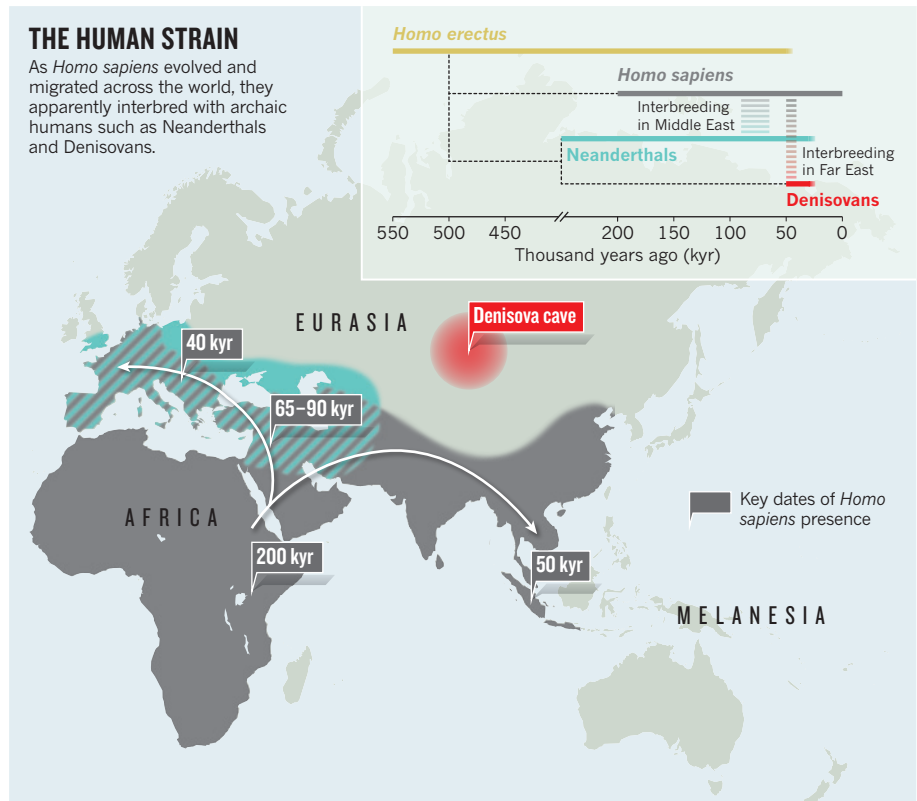
Neanderthals were large-bodied hunter-gatherers, named after the German valley where their bones were first discovered, who roamed Europe and parts of Asia from 400,000 years ago until about 30,000 years ago. The Neanderthal genome — shepherded by Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany — indicates that their evolutionary story began to split from the lineage of modern humans less than half a million years ago, when their common ancestor lived in Africa (see “The human strain”). In December last year, Pääbo’s team released the genetic blueprint of another population of ancient humans — unlike ourselves or the Neanderthals — that was based on DNA recovered from a 30,000–50,000-year-old finger bone found in a cave in Denisova in southern Siberia². Palaeoanthropologists call these groups archaic humans, distinguishing them from modern *Homo sapiens*, which emerged in Africa only around 200,000 years ago.

Pääbo is amazed at how quickly the Neanderthal genome has been mined. At a genomics meeting last year, for example, Cory McLean, a graduate student at Stanford University in California, was scheduled to talk immediately after Pääbo presented the Neanderthal genome. Inspired, McLean had trawled through the just-released genome in the days before his talk. He discovered that Neanderthals, like humans, lacked a stretch of DNA that orchestrates the growth of spines on the penises of other primates, and promptly presented the find just after Pääbo presented his³.

Since then, scientists have fleshed out the

THE HUMAN STRAIN

As *Homo sapiens* evolved and migrated across the world, they apparently interbred with archaic humans such as Neanderthals and Denisovans.



details of one of the biggest surprises from the Neanderthal genome: humans living outside Africa owe up to 4% of their DNA to Neanderthals. One explanation might be that humans migrating out of Africa mated with Neanderthals, probably resident in the Middle East, before their offspring fanned out across Europe and Asia.

By comparing individual DNA letters in multiple modern human genomes with those in the Neanderthal genome, the date of that interbreeding has now been pinned down to 65,000–90,000 years ago. Montgomery Slatkin and Anna-Sapfo Malaspinas, theoretical geneticists from the University of California, Berkeley, presented the finding at the Society for Molecular Biology and Evolution meeting in Kyoto, Japan, held on 26–30 July.

Slatkin says that their result agrees with another study presented at the meeting that came from the group of David Reich, a geneticist at Harvard Medical School in Boston, Massachusetts, who was involved in sequencing

both the Neanderthal and Denisova genomes. The dates also mesh with archaeological finds bookending early human migrations out of Africa to between about 50,000 and 100,000 years ago. Reich’s team is now developing tools to find signs of more recent interbreeding that might have occurred after humans arrived in Asia and Europe.

MORE THAN GENES

The denizens of Denisova also bred with contemporary humans, according to Pääbo and Reich’s analysis². But the only traces of their DNA to be found in modern humans were in residents of Melanesia, thousands of miles away from Denisova, suggesting that the Denisovans had once lived across Asia. In 2008, Pääbo’s team set up a lab in Beijing to screen fossils that might contain Denisovan DNA, in the hope of learning more about them and their interactions with modern humans. Currently, the bone that yielded the Denisovan genome, and a single molar from the same

cave, are their only known fossil remains, but other archaic human fossils from Asia could bear traces of this group.

Even before the Neanderthal genome made its debut in May 2010, scientists had argued that humans may have acquired not just DNA from archaic humans, but useful traits too. Human gene variants linked to brain development and speech were proposed as candidates, only to be scotched after closer inspection of the Neanderthal genome. However, a study presented at a Royal Society symposium in London in June suggests that humans owe important disease-fighting genes to Neanderthals and Denisovans. Interbreeding endowed humans with a 'hybrid vigour' that helped them colonize the world, said Peter Parham, an immunogeneticist at Stanford University School of Medicine, California, at the symposium.

Parham's team compared a group of diverse immune genes — the human leukocyte antigen (HLA) genes — in Neanderthals, Denisovans and human groups from around the world. In several cases, Neanderthals and Denisovans carried versions of HLA genes that are abundant in modern humans in parts of Europe and Asia, but less common in Africans. Varying degrees of interbreeding could explain the mismatch, Parham says. He estimates that Europeans owe 50% of variants of one class of HLA gene to interbreeding, Asians 70–80%, and Papua New Guineans up to 95%.

"It does mean that some of us owe part of our immune-system function to Neanderthals," says Pääbo. However, John Hawks, a biological anthropologist at the University of Wisconsin-Madison, notes that many HLA genes pre-date humans' split from Neanderthals and Denisovans, and that the differences may have arisen by chance as the groups evolved.

Hawks, too, has been digging into the archaic genomes, and his team has already discovered that Neanderthals and Denisovans lack certain

forms of genes that may help modern humans to fend off epidemic diseases, such as measles. This is hardly surprising: the low population density of hunter-gatherers meant that epidemics were unlikely, so they probably would not have benefited from these immune genes.

But Hawks's team is now using the find to test whether the defensive genes are linked to autoimmune diseases. In September, Hawks and his colleague Aaron Sams are scheduled to present data at a meeting of the European Society for the Study of Human Evolution in Leipzig, Germany, showing that the Denis-

"These genomes are publicly available. There's nothing stopping high-school students from doing this."

ovans lacked nearly all of the gene variants linked to coeliac disease, a gut autoimmune disorder present in modern humans. Hawks suspects that the variants may actually be in the same genes that are

linked to epidemic resistance — if they are, further study could reveal how recently such autoimmune diseases arose in humans.

Unlike most scientists mining the ancient genomes, Hawks has reported some of his more prosaic findings — Denisovans didn't have red hair, for example — on his blog (see go.nature.com/irclra). "These genomes are publicly available. There's nothing stopping high-school students from doing this, and the kind of stuff that I'm putting out on my blog is the stuff that a smart high-school student could do." More significant (and closely guarded) insights will come from developing new methods for analysing ancient genomes to test hypotheses about evolution, he says.

Pääbo, Reich and the other scientists involved in sequencing the ancient genomes are eager to see others run with their data, but caution that they need to be aware of the limitations. "They're really terrible-quality genomes", chock-full of gaps and errors and sections in which short stretches of DNA sequence have been put in the wrong place, says Reich. "There are a lot of traps in using these data, and if people are not careful they'll find all sorts of interesting things that are wrong." Pääbo's team is working on improving the quality of the sequences and including data from more Neanderthals and — he hopes — Denisovans.

Pääbo says that he and his team regularly receive e-mails from scientists asking them questions about using the ancient genomes, which they have attempted to make as user-friendly as possible. But if the first year of ancient human genomics is any indication, these requests will multiply as scientists find new applications for the genomes. "Maybe we should write a little booklet called archaic genomics for dummies," Pääbo says. ■

1. Green, R. E. *et al. Science* **328**, 710–722 (2010).
2. Reich, D. *et al. Nature* **468**, 1053–1060 (2010).
3. McLean, C. Y. *et al. Nature* **471**, 216–219 (2011).



Most of the Neanderthal genome was sequenced from bones found in Vindija cave, Croatia.

SCIENTIFIC SOCIETIES

Chinese academies promise cleaner elections

Researchers call for further reforms in selection process for prestigious membership.

BY JANE QIU IN BEIJING

Being elected as an academician in China can be a passport to political influence and financial reward. Members of the Chinese Academy of Sciences (CAS) and the Chinese Academy of Engineering (CAE) are supposed to represent the cream of the country's research, and are often invited to sit on powerful government committees. In some provinces they are offered the same official ranking as a vice-minister, and are favoured with prize money, cars, houses and personal doctors. The pursuit of these prestigious posts can involve intensive lobbying and even bribery.

This year's academy elections are already in full swing, and both societies are promising to clean up the process through transparent and rigorous oversight. They are publishing their shortlists and inviting public comment at every stage of selection, and saying that they will take allegations of misconduct more seriously than ever before. This new openness is throwing prospective academicians under increasing scrutiny.

Witness the case of Duan Zhenhao, a geologist at the CAS's Institute of Geology and Geophysics (IGG) in Beijing. He was one of 314 candidates nominated to the CAS this year; two rounds of voting will whittle that list down to a maximum of 60. But, last month, after Duan was accused on an Internet forum of embezzling research funds, the CAS began an unusually swift investigation. Duan was detained by police. Both the CAS and the IGG declined to comment on the matter, and Duan did not respond to interview requests.

Fang Shimin, a US-trained biochemist who runs a website (www.xys.org) exposing research misconduct in China, welcomes the

CAS's quick response to the allegations, but says that investigations of alleged misconduct by nominees still lack rigour and transparency.

Others say that the academy has made visible headway. "The CAS has significantly tightened up the election process in recent years," says Mu-ming Poo, a neuroscientist at the University of California, Berkeley, and director of the CAS Institute of Neuroscience

made significant contributions to engineering or clinical science become academicians. The academy has been criticized for electing a high proportion of government officials and corporate executives, and turning a blind eye to attempts to exploit the system. Nominations can come from companies and government agencies, as well as from institutions, academic associations and the academicians themselves.

These non-academic groups often run major, well-financed election campaigns, but the CAE is clamping down on such methods. In his statement to Xinhua, Xu said "those who cheat or try to use inappropriate means to get elected will lose their candidacy this year".

When the CAE announced in June that it had selected 163 candidates from a list of 485 nominees, several nominated government officials and company executives had apparently failed to make the cut — among them the deputy mayor of Shanghai and the chairman of state oil giant Sinopec. Rejected candidates had not demonstrated adequate academic contributions, says Xu.

But Gu Haibin, an economist at the Renmin University of China in Beijing, who has studied the system for ten years, believes that the academies' efforts don't go far enough. "Promises to clean up the elections mask a deeper institutional problem," he says.

Gu argues that the current academician system reinforces *guan ben wei*, the belief that political power and status are more important than any other achievement. This means that many academics act as if the ultimate goal of their academic endeavours is political advancement, rather than producing quality research.

"The whole system is rotten to the core," says Gu. "*Guan ben wei* is deeply rooted in Chinese society and is poisoning all aspects of academic life." ■



New academicians join a world of prestige and power.

in Shanghai. "I can see obvious improvement in the quality of newly elected academicians." (As a US citizen, Poo is not eligible for CAS membership.) Yao Tandong, director of the CAS Institute of Tibetan Research in Beijing, who was elected to the academy in 2007, agrees that the evaluation procedure for CAS academicians is now "probably one of the fairest systems in science in China".

Meanwhile, Xu Rigan, vice-president of the CAE, last month told the Xinhua news agency that his academy plans to improve its own election process to ensure that only those who have


**MORE
ONLINE**

TOP STORY



Antibodies linked to long-term Lyme disease symptoms
go.nature.com/ktif8o

MORE NEWS

- Gene-therapy enzymes make unpredicted errors go.nature.com/yegtcj
- Wetlands not aided by Mississippi diversions go.nature.com/qticwp
- Brazil promises 75,000 scholarships in science and technology go.nature.com/eswtrp

VIDEO



Going for a walk in the virtual countryside go.nature.com/ikedoc

ENERGY

Plutonium plans in limbo

Fukushima's economic fallout shutter's Britain's Sellafield mixed-oxide facility.

BY EDWIN CARTLIDGE

The Fukushima nuclear disaster, already casting a long shadow over the nuclear industry, has claimed another victim.

Last week, Britain's Nuclear Decommissioning Authority (NDA) announced that it will close a troubled Sellafield facility that is one of only two commercial plants in the world producing mixed oxide nuclear fuel (MOX), after reactor shutdowns in Japan eliminated its only customers for the plutonium-containing fuel. "The reason for this [closure] is directly related to the tragic events in Japan and their ongoing impact on the power markets," says Tony Fountain, the NDA's chief executive. The decision is a setback for the British nuclear industry and leaves the country wondering anew what to do with the world's largest single stockpile of plutonium.

MOX is created by combining uranium with plutonium extracted from spent reactor fuel or recycled from unused nuclear weapons. Proponents say that it provides a way to squeeze more energy out of uranium, reduce the volume of reactor waste, and destroy material that could otherwise be turned back into weapons. MOX accounts for 2% of the fuel burned in reactors around the world, 80% of it in France.

Opened in 2001, the Sellafield MOX plant was originally designed to manufacture 560 tonnes of MOX over 10 years. It would have allowed Britain to convert much of its plutonium stockpile, which now stands at 112 tonnes. This was built up over years of reprocessing spent fuel to extract plutonium, which was originally intended for use in fast-breeder reactors.

But technical failures have repeatedly halted MOX production at Sellafield. The plant has produced just 15 tonnes of MOX, at a total cost of £1.4 billion (US\$2.3 billion). Its closure leaves the Melox plant in Gard, France, operated by energy company Areva, as the world's only commercial supplier — it has produced about 1,500 tonnes of the fuel since 1995.

Last year, things were looking up for the Sellafield MOX plant. The NDA, which has owned the plant since 2005, reached an agreement with 10 Japanese electrical utilities that they would buy MOX fuel produced from Japanese plutonium stored at Sellafield. Chubu Electric Power in Nagoya was set



➔ WWW.NATURE.COM/JAPANQUAKE



C. FURLONG/GETTY

The end of MOX production at Sellafield puts Britain's huge plutonium stockpile in the spotlight.

to pay for upgrades at Sellafield's plant and buy the first consignment of fuel for its Hamaoka power station, and TEPCO, the Tokyo-based utility company that operated the shattered Fukushima Daiichi reactors, would have taken half of all the Japanese fuel produced.

But Sellafield has now lost its customers. In the wake of the nuclear disaster, only 17 of Japan's 54 nuclear reactors are currently operational, TEPCO is under serious financial and political pressure, and Hamaoka remains closed while tsunami protection is improved at the site.

As well as leading to the loss of roughly 600 jobs, the closure will intensify the debate over what to do with Britain's plutonium stockpile. Simply burying the plutonium deep underground would write off the energy that it could generate, and waste an undisclosed sum of money that has already been spent extracting it from spent fuel. But keeping it in indefinite storage is not an option, says Francis Livens, research director of the Dalton Nuclear Institute at the University of Manchester. Over time, plutonium-241 decays to americium-241, which gives off fairly intense gamma radiation that would make handling the stockpile much more difficult.

Consequently, the British government is considering building a brand new MOX plant with a much higher output, with an announcement expected in the autumn. A report produced earlier this year by the University of Oxford's Smith School of Enterprise and the Environment favoured renewed MOX production but

estimated the cost of a new plant at £2 billion to £3 billion. The government would be "crazy to consider building another MOX plant", says energy-policy analyst Steve Thomas of the University of Greenwich in London. MOX cannot be recycled indefinitely, so "all that reprocessing does is put off the day when we have to dispose of the waste, leaving the problem with people two generations from now".

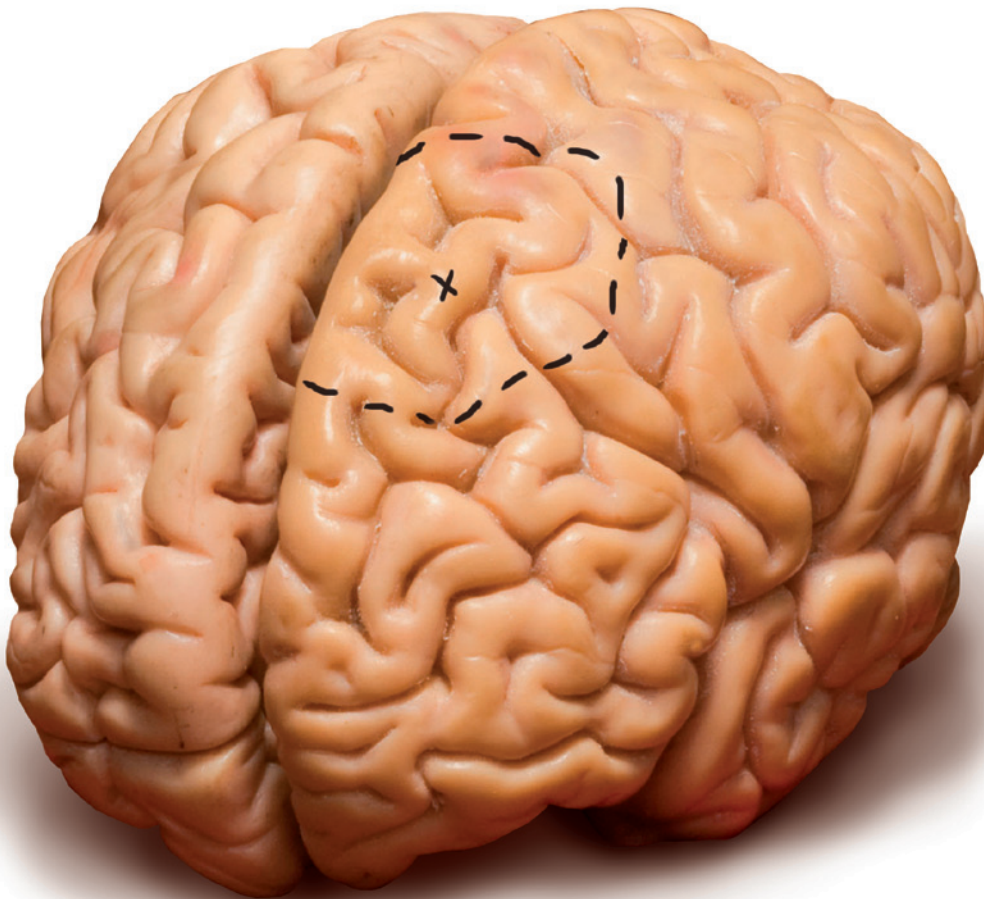
A spokesman for Areva told *Nature* that the company does not expect the Fukushima accident to significantly reduce demand for MOX in the long term. He notes that Areva is currently building the United States' first MOX plant, to recycle unwanted plutonium from nuclear weapons, at the Savannah River Site in South Carolina, and that China is also considering its own facility. "This says to us that MOX has a future," he adds.

But the Savannah River facility is hugely over budget, and will produce fuel that will cost about five times as much as conventional nuclear fuel, according to Frank von Hippel of Princeton University, New Jersey, an expert on energy policy and nuclear arms control. And although France is a stronghold of MOX production for now, he adds, its future depends on continued support for nuclear reprocessing from the French government. If the opposition Socialist Party wins next year's presidential election, Areva's position could be weakened, von Hippel argues. "Given the dominance of France in this field," he says, "that would put a question mark over the future of MOX just about everywhere." ■

WHY FAKE IT?

HOW 'SHAM' BRAIN SURGERY COULD BE
KILLING OFF VALUABLE THERAPIES FOR
PARKINSON'S DISEASE.

BY ALLA KATSNELSON



Peggy Willocks was 44 when she was diagnosed with Parkinson's disease. It progressed quickly, forcing her to retire four years later from her job as a primary-school principal in Elizabethton, Tennessee. Soon, her condition had deteriorated so much that she was often unable to dress and feed herself, take care of basic hygiene or walk unaided across a room.

Willocks enrolled in a trial for an experimental therapy called Spheramine, developed by Titan Pharmaceuticals, a biotechnology company in South San Francisco, California. Spheramine consists of cultured human retinal epithelial cells bound to specialized man-made carrier molecules. The cells are implanted into the brain, where it is hoped that they will produce the dopamine precursor levodopa, which can reduce the symptoms of Parkinson's disease. In August 2000, Willocks became the second person ever to receive the treatment. After having a steel halo — a stereotactic frame — bolted to her skull, she was put under general anaesthesia. Surgeons then used the frame and coordinates obtained from numerous magnetic resonance imaging (MRI) scans to pinpoint the location at which to drill. They then snaked a catheter through her brain's white matter to deliver the cells into the striatum.

At first there was no effect, but Willocks says that after 6–8 months she began to feel better. The changes were always moderate and gradual, except for once, about nine months after her surgery, when she showed what her doctor called a “radical” improvement in balance. By a year after the treatment, she and the five other patients in the phase I trial showed an improvement in motor ability of 48%, and those gains largely held 4 years later¹.

Ten years on, she says she notices her condition worsening, but is still doing much better than she was before her operation. She has no doubt that the treatment works. Investigators disagree: Spheramine was shelved in 2008 after a follow-up phase II, double-blind study found that it was no more effective than placebo². This time, the researchers compared the treatment with a ‘sham’ brain surgery that copied almost every aspect of the procedure Willocks received, short of injecting cells into the brain.

For many investigators aiming to treat Parkinson's and other neurological diseases invasively, using sham brain surgery as a control is, well, a no-brainer. And the practice is likely to expand in coming years, as researchers continue to develop experimental tissue transplants, gene therapies and stem-cell treatments. Small safety trials such as the one in which Willocks was enrolled may hint at the efficacy of a treatment, but they are not designed to prove it. And because they are ‘open label’ — both the investigators and the participants know that the drug is being administered — they are riddled with biases that can skew results. “It is so clear that open-label studies provide information that is not reliable,” says Warren Olanow, a neurologist at New York's Mount Sinai Medical Center who has worked on cell-based neurosurgical therapies in Parkinson's for more than two decades. “It's almost impossible for me to imagine how a serious scientist can not desire their data or hypothesis to be tested in double-blind studies.”

Other scientists, however, say that sham brain surgery is an expensive, potentially dangerous and possibly unethical bit of biomedical theatrics. It may also be unnecessary. Clinical neuroscientist Roger Barker at the University of Cambridge, UK, contends that because there is huge variation in how these therapies are administered and in how patients respond, the protocols need to be refined in an open-label setting before going on to the next stage of development. And because cost, complexity and the small number of people eligible for such invasive therapies limit the size of the studies, a sham control provides results of limited statistical utility. Barker and his colleagues across Europe are currently enrolling patients in a €12-million (US\$17-million) multicentre trial of a fetal dopaminergic nerve-cell treatment for Parkinson's disease. The treatment may never be tested against a sham-surgery control. “There's

a sort of historical precedent” for using placebo controls, but it may not apply to neurosurgical trials, he says. Willocks and other patients go further. Placebo-controlled studies aren't just unnecessary, they say, they are actually causing the downfall of potentially valuable treatments.

A COMPLICATING CONTROL

During the past 25 years, surgical therapies for Parkinson's disease have travelled a rocky road. In 1987, a report by Mexican surgeons³ described seemingly miraculous effects in two patients with severe Parkinson's who received transplants of tissue from the adrenal gland, which produces dopamine. In the next several years, hundreds of patients received the treatment, but some autopsies later showed that the cells didn't in fact survive⁴. Around the same time, researchers started to test fetal nerve-cell transplants (similar to those in Barker's trial) in small-scale studies, finding mixed but promising results. Two studies^{5,6} comparing the treatment to sham surgery concluded, however, that the transplants were not only ineffective, but also often caused dyskinesia — the movement disorder that plagues people with Parkinson's disease. In the past seven years, three experimental treatments (including Spheramine) that showed promise in small, open-label studies^{1,7,8} failed in phase II trials^{2,4,9} comparing them with a sham control (see ‘The sham wall’).

Sham brain surgery is no sugar pill. After the stereotactic frame is affixed to the skull, the patient is usually anaesthetized and surgeons drill into the skull. In most cases, the burr holes stop at the dura mater, a protective membrane covering the brain, but they sometimes go deeper: in a phase II trial testing the nerve growth factor GDNF, investigators catheterized the brain in all participants but infused saline, rather than GDNF, into the controls⁹.

“We have to stage the whole thing such that from the outside it's completely indistinguishable” from the real thing, says Joao Siffert, chief medical officer of Ceregene, a company in San Diego, California, that is working on a therapy that delivers a gene for another nerve growth factor, called neurturin, using a viral

vector. For many sham treatments, everyone in the operating room, from surgeons to nurses' assistants, must pretend that they are busily performing the complete operation — in some cases, turning on machines to elicit appropriate noises. An extremely complex protocol ensures that no one outside the surgical team knows who got what treatment. “It's very complicated, there are a lot of moving parts,” Siffert says. All that ratchets up the cost of a trial; Siffert estimates that between operating-theatre costs, follow-up and the unwieldy infrastructure required for data management, a 50-patient study would cost more than \$10 million.

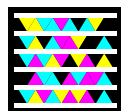
Still, at least in North America, Parkinson's disease investigators overwhelmingly support the use of sham surgery — at a rate of 94%, according to a 2004 survey¹⁰. Around 20% said that penetrating the brain is justifiable. And proponents say the procedure is relatively safe. Although sham brain surgery has definite risks, most notably those associated with general anaesthesia, supporters note that adverse events are almost unheard of, unlike the risks of the actual treatments. And participants in the sham groups are generally promised the treatment if it is ultimately approved; in that event, they will already have the burr holes in their skulls through which it would be administered.

Sham treatments help to tease out the placebo effect and biases. In Parkinson's disease, the placebo effect is especially strong. One reason is that patients' expectations that they will benefit from a treatment induce the release of dopamine¹¹, the neurotransmitter that is lacking in the disease. “The placebo effect is real, it's huge and it's got a physiological basis,” says Jon Stoessl, a neurologist at the University of British Columbia in Vancouver,

THERE'S A HISTORICAL PRECEDENT FOR USING PLACEBO CONTROLS, BUT IT MAY NOT APPLY TO NEUROSURGICAL TRIALS.

NATURE.COM

To listen to a podcast on sham surgery: go.nature.com/8seayr



Scan the tag above with the free app from gettag.mobi

THE SHAM WALL

In the past decade, sham-controlled studies have dashed hopes for all but a few surgical treatments for Parkinson's disease.

Therapy	Open label (phase I)	Sham controlled (phase II)
Transfer of gene for glutamic acid decarboxylase (GAD)	12 patients; 19% improvement at 3 months that persisted at 1 year ¹⁴	45 patients; 23% improvement in treatment group vs 13% in controls at 6 months ¹⁵
Implantation of cultured retinal cells (Spheramine)	6 patients; 48% improvement at 1 year, 43% improvement at 4 years ¹	71 patients; no significant difference after 1 year ²
Transfer of gene for neurturin	12 patients; 36% improvement at 1 year ⁷	58 patients; no significant difference after 1 year ⁴ , but 18-month follow-up and autopsy data spurred a second trial, now ongoing
Infusion of glial-cell-line-derived neurotrophic factor (GDNF)	5 patients; 39% improvement after 1 year ⁸	34 patients; no significant difference after 6 months ⁹

Improvements are for off-medication unified Parkinson's disease rating scale (UPDRS) motor score.

Canada, who studies Parkinson's and the placebo effect. In one double-blind study of fetal nerve-cell transplants, patient improvement correlated with whether they believed they had received the treatment, irrespective of whether they actually had¹². And the effect can last as long as two years, Stoessl says, citing an unpublished study by his colleagues.

Many regard bias as a more significant confounder. "Investigators have a tremendous vested interest in seeing that their treatment is effective," says Anthony Lang, a neurologist at the University of Toronto in Canada who has participated in several neurosurgical trials for experimental Parkinson's therapies. In any trial, bias can affect how researchers assess patient responses and may inflate the patients' expectations, further enhancing the placebo effect. Compounding the problem for Parkinson's research is the fact that there are no objective measures for how well a patient is doing. "It's just a sort of perfect storm conspiring against our ability to see definitive changes in the underlying disease," says Steven Piantadosi, a clinical-trials methodologist at Cedars-Sinai Medical Center in Los Angeles, California. "Sham surgery, properly done, can control for that."

Barker counters that it is possible to control for investigator bias in an open-label trial by taking steps such as having blinded raters assess patients. His position is in some ways unsurprising; in Europe, sham surgery is deemed much less acceptable than it is in the United States. It has never been used in the United Kingdom. Barker is categorical about his belief that transplantation of fetal tissue works, at least for some people. "I don't need sham surgery to show that," he says, pointing instead to a paper¹³ published last year describing two patients treated 13 and 16 years previously who were still benefiting from the treatment, and whose brains showed functional dopamine-producing neurons at the transplant site. He attributes the mixed results in past studies to variation in the patients selected for treatment, the characteristics of the tissue being implanted and the methods used to implant it. His trial will have to demonstrate efficacy without eliciting some of the side effects found in the two sham-controlled studies. That will require some type of control study, he says, but it might take the form of comparison to an approved therapy that is known to work, such as deep brain stimulation.

But time, says Barker, will best establish efficacy. In most trials the end point is no more than a year after the treatment. That may not be long enough: implanted cells or injected growth factors might take longer than this to become fully functional, and the placebo effect may not have had time to dissipate. "We want a 3–5-year endpoint," says Barker.

There are hints from some of the failed phase II trials that patients followed up beyond study endpoints might tell a more positive story⁴. Some say, therefore, that sham controls are sinking the prospects of valuable drugs. Anders Björklund, a neuroscientist at Lund University in Sweden who is collaborating with Barker, says that sham surgery can lead researchers to throw out a strategy prematurely if the trial fails because of technical or methodological glitches rather than a true lack of efficacy.

ADVOCACY AND FRUSTRATION

According to Perry Cohen, who leads a network of patient activists called the Parkinson Pipeline Project, that's exactly what is happening. He had always questioned the need for sham surgery, he says, but after the string of phase II failures, "We started saying, 'Hey, this is a problem. These trials failed, but we know they are working for some people.'"

For researchers, it is easy to dismiss patients' concerns as being driven by emotion. "Patients want cures," says Lang, "and they will often be convinced that the more aggressive, surgical therapies are more likely to be curative." But Cohen counters that patients have different priorities and that researchers must take these into account. Researchers use placebo controls to weed out false positives. But for patients, the real ogre is the false negatives — which can sink a therapy before it has been optimized. The better a trial is at stamping out the former, the higher the rate of the latter — which means at best delays, and at worst dead ends. Spheramine, for example, "is still on a shelf somewhere," Cohen says. Then there's Amgen's phase II trial of GDNF. The trial was halted in 2004 amid lacklustre results and potential safety concerns, which some have attributed to Amgen's procedure, rather than to the therapy itself. Now researchers are taking a renewed interest in the molecule, but although Cohen is glad it is getting a second chance, "we lost 6 years on it," he says.

Patients also have different perspectives on risk from researchers, Cohen says. He offers the story of Tom Intili, who had had Parkinson's for 10 years when, at the age of 50, he signed on to the double-blind, placebo-controlled trial of neurturin. At first, Intili improved dramatically. But when the results were unblinded, he learned that he had received the sham. His condition plummeted, leaving him more debilitated than he had been before the trial. "We just don't know what the psychological effects of unblinding are," Cohen says.

Moreover, trying to exclude the placebo effect is simply misguided, Cohen argues. "I don't want to subtract out the placebo effect — I want to keep it, because in real life it's part of the treatment," he insists. Because psychological factors are so salient in Parkinson's, a placebo response might actually potentiate a therapy, he explains. "I want to be convinced that sham surgery is necessary. I'm looking for arguments that might change my mind, but I haven't found any yet," he says.

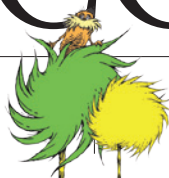
Willocks says that she is living proof that many of the recently shelved therapies are in fact salvageable. Of course from a scientific perspective, her story is an anecdote, not data. In May, the failed phase II study of Spheramine — the therapy she received a decade ago — was finally published². The paper closes with a warning about the dangers of the placebo effect and stresses the importance of controlling for it with a double-blind design. "That last paragraph bothered me," Willocks says. "I just don't see how they can call it a placebo effect after ten years." ■

Alla Katsnelson is a freelance science writer in New York City.

1. Stover, N. P. & Watts, R. L. *Neurotherapeutics* **5**, 252–259 (2008).
2. Gross, R. E. *et al. Lancet Neurol.* **10**, 509–519 (2011).
3. Madrazo, I. *et al. N. Engl. J. Med.* **316**, 831–834 (1987).
4. Marks, W. J. *et al. Lancet Neurol.* **9**, 1164–1172 (2010).
5. Freed, C. R. *et al. N. Engl. J. Med.* **344**, 710–719 (2001).
6. Olanow, C. W. *et al. Ann. Neurol.* **54**, 403–414 (2003).
7. Marks, W. J. *et al. Lancet Neurol.* **7**, 400–408 (2008).
8. Gill, S. S. *et al. Nature Med.* **9**, 589–595 (2003).
9. Lang, A. E. *et al. Ann. Neurol.* **59**, 459–466 (2006).
10. Kim, S. Y. H. *et al. Arch. Neurol.* **62**, 1357–1360 (2005).
11. de la Fuente-Fernández, R. *et al. Science* **293**, 1164–1166 (2001).
12. McRae, C. *et al. Arch. Gen. Psychiatry* **61**, 412–420 (2004).
13. Politis, M. *et al. Sci. Transl. Med.* **2**, 38ra46 (2010).
14. Kaplitt, M. G. *et al. Lancet* **369**, 2097–2105 (2007).
15. LeWitt, P. A. *et al. Lancet Neurol.* **10**, 309–319 (2011).

COMMENT

ENVIRONMENT Classic children's eco fable turns 40 **p.148**



HISTORY The travails of the first mission to calibrate latitude **p.149**

EXHIBITION Images of the mind, from Golgi to Gormley on show in Dresden **p.151**

POLICY Defending US recommendations for lab animal use and care **p.152**

HARPERCOLLINS



The arXiv server in the early 1990s: a computer that helped to change the world of physics.

ArXiv at 20

Paul Ginsparg, founder of the preprint server, reflects on two decades of sharing results rapidly online – and on the future of scholarly communication.

Twenty years ago this month, I launched an electronic bulletin board intended to serve a few hundred friends and colleagues working in a subfield of theoretical high-energy physics. I had recently moved to the Los Alamos National Laboratory in New Mexico and for the first time had my own computer on my desk, and the desire to simplify the exchange of unpublished manuscripts (preprints) between researchers, previously distributed as paper copies by post.

This automated repository and alert system for physics preprints, at hep-th@xxx.lanl.gov, was implemented shortly before

the dawn of the web era. As I e-mailed to a colleague at CERN more than a year later: 'I know nothing of WWW, what is it?' The original plan was for roughly 100 full-text article submissions every year, each stored for three months until the existing paper distribution system could catch up. By popular demand, nothing was ever deleted.

Within a few years it had evolved into a web resource at arXiv.org that now contains close to 700,000 full texts, receives 75,000 new texts each year, and serves roughly 1 million full-text downloads to about 400,000 distinct users every week (see graphs). It has broadened, first to cover most active research

fields of physics, then to mathematics, nonlinear sciences, computer science, statistics and, more recently, to host parts of biology and finance infiltrated by physicists.

It is heartening, 20 years later, to see a stable and successful arXiv, running some of the original software and providing services to a community nearly a thousand times larger than expected. But at some point a thorough overhaul will be needed to keep pace with new online trends and opportunities.

For me, the repository was supposed to be a three-hour tour, not a life sentence. ArXiv was originally conceived to be fully ►

J. FLOWER/LANL

► automated, so as not to scuttle my research career. But daily administrative activities associated with running it can consume hours of every weekday, year-round without holiday. So, from September, the site will be entirely in the hands of the staff of Cornell University Library in Ithaca, New York. I will remain on the advisory board and continue some research projects in text and data mining and in supporting next-generation document formats and information filters.

To reflect on ArXiv is to reflect on a research world transformed by a two-decade revolution in information technology, with vast quantities of literature and associated resources now available on demand. Yet, it is a surprise that scholarly publishing as a whole remains in transition. There is no consensus on the best way to implement quality control (top-down or crowd-sourced, or at what stage), how to fund it or how to integrate data and other tools needed for scientific reproducibility.

My hope is that rather than merely using electronic infrastructure as a more efficient means of distribution, the revolution-in-waiting will ultimately lead to a more powerful knowledge structure, fundamentally transforming the ways in which we process and organize scientific data.

DEMOCRATIC GOALS

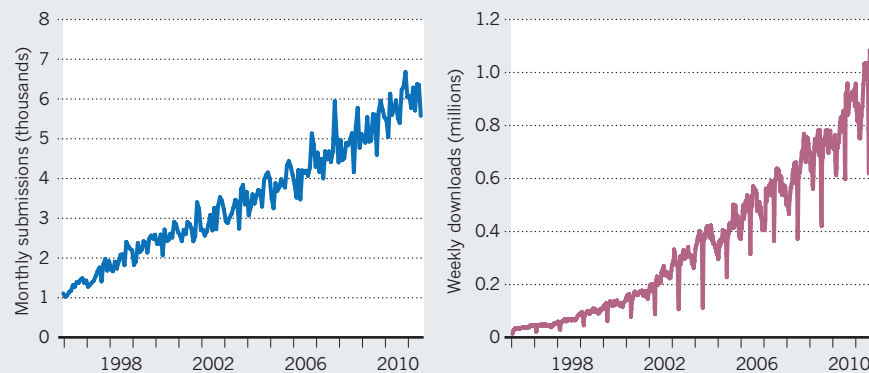
The original bulletin board was engineered to level the research playing field. It is hard to imagine now, but considerable time and effort was once spent printing, photocopying and posting preprints to a privileged few friends and colleagues, before publication in formal journals. The idea of a central repository was to allow any researcher worldwide with network access to submit and read full-text articles, giving equal entry to everyone from graduate students up. (The early Internet was an academic playground — the general public didn't start coming online until a few years later.)

Within two years, arXiv had evolved into the primary daily resource for a global community of researchers. It became a place to stake intellectual precedence claims, catalysing further growth.

Launched in 1991, before any conventional journals were online, arXiv pioneered many of the tools now taken for granted. We led the way in using the abstract page as a hub to diverse formats and resources, linking author names automatically to search functions, and we were early adopters of electronic formats from compressed postscript to PDF for file sharing. The repository showed that researchers were willing and eager to move to fully electronic means of dissemination. It also foreshadowed the 'interactive web', insofar as it provided a rudimentary framework for users to deposit content. The value of this

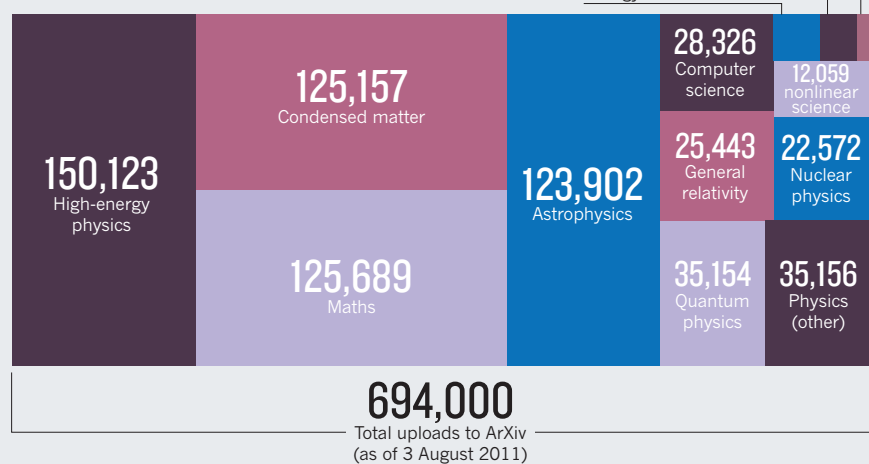
DIGITAL PIONEERS LEAD THE WAY TO SHARING RESEARCH ONLINE

The popularity of the arXiv preprint server has grown inexorably since its launch in the early 1990s. Academics enjoy the universal access, low cost and speed of online distribution.



PHYSICS ENVY

Mathematicians, astrophysicists and even some biologists have joined high-energy physicists in uploading articles to ArXiv.



content is then amplified by sharing.

In many ways, building the technology was simpler than managing the sociological and financial aspects. A decade ago, the main site moved with me and became embedded within Cornell University library. Although it serves more users on a daily basis than any other library resource, most of those users are external, so it is less clear where it should fit in university funding priorities. In the absence of a wealthy donor willing to provide a small endowment in exchange for far more name recognition than any traditional building donation (hint, hint), the library has recently asked institutions who are heavy users to contribute to running costs (<http://arxiv.org/help/support>). This distributes the financial burden and oversight to a larger community, while buying time to investigate long-term business models.

Physicists were quick to adopt widespread sharing of electronic preprints, but other researchers remain reluctant to do so. Fields vary widely in their attitudes to data and ideas before formal review, and in choosing to share electronic preprints, each

community will have to develop policies and protocols best suited to their users. A talk I gave in 1997 to a group of biologists helped catalyse the resource now known as PubMedCentral — run by the US National Institutes of Health. I served on the initial advisory board, which soon decided not to host any unrefereed materials, even carefully quarantined, in part for fear of losing essential publisher participation. There remain many legitimate reasons for individual researchers to prefer to delay dissemination, from uncertainty over correctness, to retaining extra time for follow-ups, to sociological differences in the way publication is regarded — in certain fields, the research somehow doesn't count until peer reviewed.

No community that has adopted arXiv usage has renounced it, however, so the growth has been inexorable. Adoption by some fields, including computer science, started as a trickle, increasing dramatically many years later. Even some subfields of physics have experienced delayed adoption: emerging research in 2008 into superconductivity in iron compounds brought in a group

of condensed-matter experimentalists traditionally more cautious about disseminating results. They were ultimately won over by the need to stake precedence claims and get their results in front of theorists.

Even today, fields vary hugely in how they recognize intellectual precedence. It baffles me that scientists in some fields can announce a result in a public forum, such as a meeting, while another group can reproduce the results, publish first in a journal, and be given complete intellectual precedence, as though the information did not exist until vetted by the referee process. Journal editors and referees should make more effort to ensure proper attribution is given to publicly accessible materials in a stable resource, such as arXiv.

WHERE NEXT?

As the arXiv community has diversified, so have its desires. Some users have requested support for comment threads related directly to papers on the site, while others prefer that it maintain its unadulterated stream of author-provided content. I have sympathy for more interactivity: in today's social web, a one-way channel seems an anachronism. But because maintaining utility and civility in online discussions can be labour-intensive, our policy has been that such services should be external to the main repository. The same considerations apply to self-organized, or 'crowd-sourced' forms of review. In recent years, some external blogs have begun to host useful comment threads, linked back from arXiv abstract pages through a track-back mechanism, but I don't predict that dedicated blogging by individual scientists will grow substantially. More scalable tools for discussion are provided by question-and-answer sites such as mathoverflow.net, where expert mathematicians, in the course of answering one another's posted questions, provide links to maths articles hosted on arXiv.

Again, because of cost and labour overheads, arXiv would not be able to implement conventional peer review. Even the minimal filtering of incoming preprints to maintain basic quality control involves significant daily administrative activity. Incoming abstracts are given a cursory glance by volunteer external moderators for appropriateness to their subject areas; and various automated filters, including a text classifier, flag problem submissions. Although the overall rate of such submissions is well below 1%, they tend to cluster in specific areas (such as general relativity, quantum mechanics and unified theories in

physics; proofs of the Riemann hypothesis, Goldbach's conjecture and new proofs of Fermat's last theorem

in mathematics; P versus NP problem in computer science).

Moderators, tasked with determining what is of potential interest to their communities, are sometimes forced to ascertain 'what is science?' At this point arXiv unintentionally becomes an accrediting agency for researchers, much as the Science Citation Index became an accrediting agency for journals, by formulating criteria for their inclusion. Although decisions are biased towards permissiveness, inevitably some authors object that it is never permissive enough.

DIGITAL GENERATION

The idea that print journals had outlived their usefulness was already in the air in the early 1990s. David Mermin memorably wrote in *Physics Today* in 1991: "The time is overdue to abolish journals and reorganize the way we do business."¹ By the mid 1990s, it seemed unthinkable that free and unfettered access to non-refereed papers on arXiv would continue to coexist indefinitely with quality-controlled but subscription-based publications. Fifteen years on, researchers continue to access both, successfully compartmentalizing their different roles in scholarly communication and reward structures.

The transition to article formats and features better suited to modern technology than to print on paper has also been surprisingly slow. Page markup formats, such as PDF, have only grudgingly given way to XML-based ones that support features such as manipulable graphics, dynamic views, linked annotations and semantic markup. Part of this caution is a result of the understandable need to maintain a stable archive of research literature, as provided by paper over centuries.

"It is a surprise that scholarly publishing as a whole remains in transition."

Configuring scholarly communication infrastructure for the next generation of researchers requires getting into the heads of current undergraduates and graduate students. Their life experience is of immediate online availability and global search engines, and they arrive imbued with the social-network mentality of sharing links, photos, videos and status updates. Yet, my own informal survey of graduate students reveals information-gathering techniques familiar to most older scientists. Students still follow citation trees, search by keywords and consult with peers and mentors, with the latter as important as ever for weeding out unreliable sources. Students also say that they search preferentially for open-access resources when working from home, because accessing subscription-based journals, even when available through an institutional proxy, can be frustratingly painful.

Navigating increasing quantities of data inevitably raises concerns of information overload. This phenomenon is documented back to the dawn of writing, accelerated with the invention of the printing press, and has been re-emphasized by every generation since. The superficial response is a call for better filters, but an imperfect filter can be more harmful than none. For example, commonly used recommender systems based on passive measures of global popularity can broaden individual reading choices, but effectively broaden everyone in the same direction, thereby leading to less overall community diversity. (The cynic would say reinforcing faddishness in already faddish fields.)

On arXiv, we have seen some of the unintended effects of an entire global research community ingesting the same information from the same interface on a daily basis. The order in which new preprint submissions are displayed in the daily alert, if only for a single day, strongly affects the readership on that day and leaves a measurable trace in the citation record fully six years later^{2,3}. Some researchers, wise to this, time their submissions to arrive just after the daily afternoon deadline to maximize their prominence in the next day's mailing. Filters that highlighted 'popular' materials over longer periods of time would exacerbate this effect. Hence any recommender system on arXiv would need, at minimum, to be personalized to individual readership preferences and interests to reduce herding behaviour. Experiments with such systems are ongoing, and may be put online within a year or two if they perform properly.

For now, the open questions of arXiv's long-term role and its relationship to conventional publishing, the details of its funding model, and its overall intellectual supervision, are to be resolved in coordination with its users and stakeholders. A meeting of international sponsoring institutions will be hosted by the Cornell Library next month to discuss the transition of arXiv to a collaboratively governed, community-supported resource. It will be a challenge to keep it attuned to the needs of future generations of researchers.

My hope is that the barrier to implementation of new ideas in this realm will remain low enough that, if all else fails, some young researcher elsewhere can launch another tiny ship on a fateful trip. ■

Paul Ginsparg is at the Department of Physics, Cornell University, Ithaca, New York 14853, USA.
e-mail: ginsparg@cornell.edu

1. Mermin, N. D. *Phys. Today* **44**, 9 (1991).
2. Haque, A. & Ginsparg, P. J. *Am. Soc. Inf. Sci. Technol.* **60**, 2203–2218 (2009).
3. Haque, A. & Ginsparg, P. J. *Am. Soc. Inf. Sci. Technol.* **61**, 2381–2388 (2010).

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/m3u6oj



HARPERCOLLINS CHILDREN'S BOOKS

The Lorax leaves in despair after failing to save his ecosystem from poisoning by the Thneed factory, run by the greedy, green-armed Once-ler.

IN RETROSPECT

The Lorax

Emma Marris reflects on a classic children's fable that still has lessons for environmental policy 40 years on.

Theodor Seuss Geisel, better known as Dr Seuss, wrote more than 40 children's books, beloved for their zany rhymes and sinuous illustrations. In August 1971 — the year after the creation of the US Environmental Protection Agency and celebration of the first Earth Day — Seuss published a book that became a kind of *Silent Spring* for the playground set.

Thousands of children have learned about environmental destruction from the *The Lorax*, Seuss's tale of ecological ruin brought on by greed. The book still resonates: Universal Studios is due to release a feature-length animation of it next year. It packs in a lot of sophisticated concepts for a picture book, from the interconnectedness of ecosystems to the effects of industrial pollutants on freshwater systems. There is even a trophic cascade — a shift in top predators that triggers changes through a food chain.

And what initially seems like a simplistic take on environmental policy — industry bad, activists good — turns out to be more subtle. The hero does not save the day; that task falls to the next generation. This downbeat, if realistic, plot arc makes me hesitant about introducing the book to my young daughter.

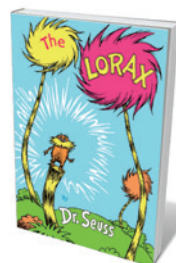
An ecologist might classify the book's lost paradise as a 'Truffula savannah'. The key-stone species are the Truffula trees, which look like candy-coloured palms. In the story, every last one is chopped down by a faceless entity, the Once-ler, to provide the raw material for a multipurpose garment called a Thneed — anticipating Snuggies and Slankets by some four decades.

Clearing the Truffula trees sets off a chain reaction that demonstrates the interdependence of life. Without Truffula fruits, the ursine Brown Bar-ba-loots have nothing to eat.

This being a children's book, they don't go extinct. Rather, they are packed off, with much lamentation, to points unknown by the creature who gives the book its title: the Lorax, a diminutive, grandly moustached character who acts as advocate for the ecosystem's species.

The Lorax also complains about the unregulated Thneed factory, which belches out smog and dumps into a pond an astonishing quantity of industrial by-products known as Gluppity-Glupp and Schloppity-Schlopp. (According to the US National Resources Defense Council, textile factories pollute 200 tonnes of water per tonne of fabric produced.) The smog chases off the ecosystem's avian endemics, the Swomee Swans, and the polluted water gums the gills of the Humming Fish, which must traipse away in search of cleaner ponds. Eventually, the bright Truffula savannah is replaced by a lifeless wasteland.

Naturally, the Once-ler gets his comeuppance. By harvesting too many Truffula trees too quickly, he put himself out of business and retreats to his ruined factory



The Lorax
DR SEUSS
First published by
Random House: 1971.
Republished by
HarperCollins
Children's Books:
2009. 64 pp. £4.99

to ruminate on the costs of not having a sustainable business plan.

The Lorax leaves in despair, and the Once-ler hands over the task of restoring the Truffula ecosystem by giving the world's last Truffula seed to a child.

The Lorax himself is a parody of a misanthropic ecologist: "He was shortish. And oldish. And brownish. And mossy. And he spoke with a voice that was sharpish and bossy." He hectors and pleads, he "speaks for the trees", but the Once-ler pays no attention until it is too late, saying, "All you do is yap-yap and say, 'Bad! Bad! Bad! Bad!'" That approach seems a little dated in this era of 'win-win' solutions. But Seuss understood, even back then, the limits of gloom and doom. The Lorax fails. Nevertheless, Seuss clearly had great affection for his impassioned little nag, as do the book's legions of fans.

Perhaps the source of its enduring appeal has less to do with Seuss's prescience about the futility of ecological doomsaying than with environmentalists' nostalgia for a time when such problems seemed more black and white. *The Lorax* portrays a world without the complexities of carbon trading, the pricing of ecosystem services, uncertainties over baseline states and the existence or not of pristine wilderness. For Seuss, wild nature is a paradise, industry is a malignant cancer and heroes take a stand.

Then again, that old narrative has not been entirely displaced by the bureaucrats who set green targets at international meetings. Diverse natural land is threatened by industries; smog chokes skies; Gluppity-Glupp fouls waterways; people use natural resources like there is no tomorrow. We still need heroes to speak for the trees.

Will I read my daughter *The Lorax* when she graduates from *Pat the Bunny*? I'm not sure. It captures the basics of why mismanagement and overexploitation of ecosystems are a bad idea in a way that children can understand. And because I believe that her generation's challenge will be to manage the planet consciously, perhaps it is not too early to teach those lessons. But this is a gloomy book. The final image — of the Truffula seed hurtling into a tiny pair of hands — puts a lot of responsibility on small shoulders. Perhaps it is more important for her to learn the pleasures and beauty of nature first. Afternoons spent poking in the mud, catching cicadas and mapping out squirrel routes may be more likely to turn her green than the spectre of a world without Truffula trees. ■

Emma Marris writes for *Nature* from Columbia, Missouri.

HISTORY

How Earth shaped up

Andrew Robinson enjoys an account of the first expedition to the equator to calibrate latitude.

What was the world's first international scientific expedition? Scientist-explorers of the Enlightenment, such as Alexander von Humboldt, would have answered without hesitation: the French government's eight-year geodesic mission to the Spanish colony of Peru, beginning in 1735. The mission — the subject of *Measure of the Earth* — calculated the precise length of a degree of latitude at the equator, thereby helping to define Earth's shape.

Today, the geodesic mission's fame has faded. Until this book, no modern history of the mission has been published in English (unpublished academic studies exist). It was virtually ignored in Ken Alder's prize-winning account of the surveying of the arc of the meridian and the origins of the metric system, *The Measure of All Things* (Free Press, 2002). Neil Safier's *Measuring the New World* (University of Chicago Press, 2008) focused instead on the Amazon explorations of its best-known member, Charles-Marie de La Condamine. But in 2006, the French geodesic mission was the basis for one episode in a BBC television series, *Voyages of Discovery*, for which the historical consultant was science writer and naval historian Larrie Ferreiro. In *Measure of the Earth*, Ferreiro puts this important, dramatic and gruelling trip back on the map.

► NATURE.COM

For more on measuring the meridian:
go.nature.com/xrgvrs



Measure of the Earth: The Enlightenment Expedition That Reshaped Our World

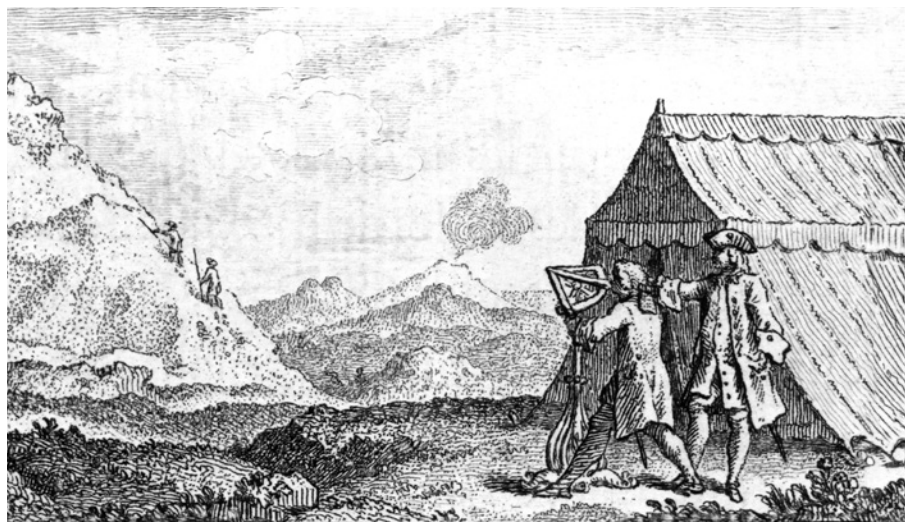
LARRIE D. FERREIRO
Basic Books: 2011.
376 pp. \$28, £15.99

The geodesic mission was organized by the French Academy of Sciences, under the sponsorship of the French minister of the navy acting in wary collaboration with his Spanish counterpart in Madrid. Spain sent two naval officers, also astronomers, to learn from and keep an eye on the French.

Along with a similar French scientific mission to the Arctic Circle in 1736–37, the expedition deter-

mined that a degree of latitude is shorter at the equator (where it measures a length of 56,753 toises, or 110.54 kilometres) than at Paris (57,060 toises) or towards the North Pole (57,437 toises). Thus it proved that Earth was not a sphere but an oblate spheroid: it bulges slightly at the equator and is slightly flattened at the poles.

This asymmetry explains why the force of gravity is slightly less near the equator than in Paris. This was first observed by a puzzled French astronomer measuring the period of a pendulum clock in Guyana in 1672, and seized on by Isaac Newton in his *Principia Mathematica* of 1687. Newton predicted the oblate shape of Earth from his theory of



The 1730s French geodesic mission revealed that Earth bulges at the equator.

ROYAL ASTRON. SOC./SPL

gravity. However, his idea was strongly contested in France by followers of philosopher René Descartes, whose theory of vortices in the ether had led them to the view that Earth must be prolate, like an upright egg. The geodesic mission thus settled a long scientific debate, and at the same time improved the accuracy of navigation by ships.

By 1735, the basic geometric and astronomical principles of trigonometric surveying were well understood. To the government and scientists in Paris, it seemed straightforward to apply these accurately among the peaks of the Andes, over a period of a year or two. But the expedition members had reckoned without a

"The expedition had reckoned without obstacles including yellow fever, local love affairs and brigandry."

myriad of obstacles, large and small, including precarious mountain transport, extreme weather, earthquakes, altitude sickness, yellow fever, local love affairs and brigandry. They suffered the theft of their triangulation signals by local Indians, damage to instruments, the withdrawal of their French funding, Spanish colonial politics and the ripples from European wars, as well as the murder of their surgeon at a bullfight.

There were also purely scientific challenges, such as stellar aberration: the then little-understood phenomenon of tiny variations in the location of the stars used to fix latitude, which we now know arises as a result of Earth's motion around the Sun.

The most serious problem was the expedition's lack of leadership. Internecine squabbling nearly scuppered the project. The precious results announced to the French Academy in 1744 by expedition member and geophysicist Pierre Bouguer were those measured by himself and La Condamine, without the knowledge of the official expedition leader, Louis Godin, who decided to remain in South America as a professor of mathematics in Lima. *Measure of the Earth* accomplishes its mission with skill and devotion, although it lacks some necessary diagrams. Its intermixing of politics and science is particularly fascinating. It also shows, unintentionally, the astonishing development of international scientific cooperation since the days of the Enlightenment pioneers. ■

Andrew Robinson is author of *The Story of Measurement*.
e-mail: andrew.robinson33@virgin.net

ANTHROPOLOGY

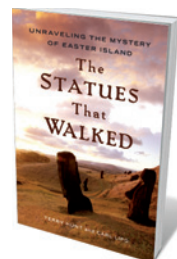
Head to head

A scenario blaming rats for the devastation of Easter Island doesn't account for recent results, argues **Paul Bahn**.

Easter Island, or Rapa Nui, is a perennial favourite of scholars and the media because of its numerous giant stone figures and supposed mysteries. Most of these enigmas — including the origins of the statues and the denuded landscape — have been solved in recent decades through painstaking work by archaeologists, anthropologists, environmentalists, linguists and geneticists.

A consensus view has emerged, summarized in *Easter Island, Earth Island* (Thames and Hudson; 1992) by botanist John Flenley and myself, that the island was deforested by its inhabitants. They cleared the land for crops and used timber for the transport and erection of ever more statues, with war the end result. In *The Statues That Walked*, archaeologists Terry Hunt and Carl Lipo present a different picture, portraying the islanders as environmentally sensitive and peace-loving until Europeans arrived in the eighteenth century. But some recent publications don't support their hypothesis.

Since the first known visit to the island by European vessels in 1722, people have wondered how so many huge stone statues could be transported and raised, given the lack of available timber. Discoveries of root moulds, pollen grains and stumps showed that the island was originally covered in millions of large palms and other tree species. Analyses of pollen and plant macrofossils revealed drastic deforestation between the thirteenth and



The Statues That Walked: Unraveling the Mystery of Easter Island

TERRY HUNT AND CARL LIPO
Free Press: 2011.
256 pp. \$26

seventeenth centuries, long before the Europeans arrived.

A wide range of evidence, along with oral traditions, suggest that the Easter Islanders had lived cooperatively for centuries after their arrival from Polynesia, probably in the early centuries AD. A thousand years later, they were in conflict, living in a barren landscape. The ecological decline of this small island serves as a warning

of what is happening to Earth as a whole, as argued by Flenley and myself, and by geographer Jared Diamond in his best-seller *Collapse* (Viking, 2004).

Inevitably, this scenario has elicited doubts. Archaeologists Catherine and Michel Orliac have speculated that the island's deforestation was largely due to drought or climatic change, which may indeed have played a part. Others, including Hunt and Lipo, contend that the islanders' ills were caused by the arrival of Europeans, rather than internal social pressures.

In *The Statues That Walked*, Hunt and Lipo argue for a late date of around 1200 AD for the



Debate surrounds the transportation of Easter Island's statues and the disappearance of its forests.

PHOTOLIBRARY/CORBIS

islanders' appearance on Rapa Nui. They claim that deforestation was mainly caused by the rats that came with them. Having found some palm nuts bearing gnaw-marks, they attribute the extinction of the island's big palm to rat predation, although they say little about other tree species. The statues, they posit, were moved upright for many miles by swivelling, which required little timber. And despite the deforestation, they say, the islanders continued to grow sufficient food and remain free of quarrels until Europeans brought violence, germs and eventual devastation.

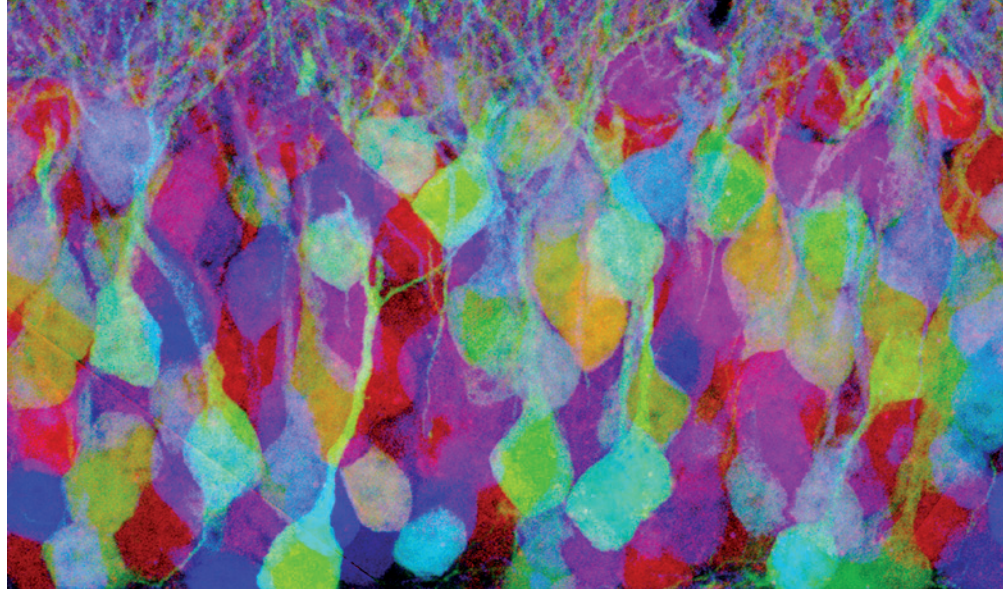
The authors' new scenario does support the consensus view that the island was only colonized once, by Polynesians — and not, contrary to the theory of Norwegian adventurer Thor Heyerdahl, by Amerindians from the New World. The book contains good passages on the carving and transportation of the statues — which we know from oral testimony to be ancestor figures that were venerated by the islanders — including an account of the pioneering excavations of the statue roads by US geologist Charlie Love. The practice of lithic mulching, in which millions of stones are spread over the soil to retain moisture for crops, is described in detail.

But coverage of work by others is incomplete. For instance, the authors mention only their own survey of the statues and not the decades-long (and ongoing) cataloguing by US archaeologist Jo Anne Van Tilburg. Nor do they note some recent published evidence that, in my view, refutes the book's basic tenets.

For example, a variety of evidence contradicts their claim of rat predation: numerous palm fruits not gnawed by rats, palm stumps burned and cut, continued germination of palms despite the rats' presence, and the disappearance of other plant species that coexist with rats elsewhere. Hunt and Lipo's claim that human skeletal remains show little evidence of lethal trauma is refuted by quotes from anthropologist Douglas Owsley, the author of a 1994 paper that they reference. After examining more than 600 Easter Island skeletons, Owsley stated in a 2003 BBC documentary that the extreme frequency of injuries proved that these were people at war: "They're slugging it out, there's no doubt about it."

Hunt and Lipo present some of the island's many features entertainingly, but the history of Rapa Nui is more complex than they allow. ■

Paul Bahn is the author (with John Flenley) of *Easter Island, Earth Island. An updated third edition is out this month.* e-mail: pgbahn@anlabryd.karoo.co.uk



The Brainbow technique colours individual neurons, here in a slice of mouse hippocampus.

NEUROSCIENCE

Picturing the soul

Alison Abbott revels in a stellar mix of brain imageries.

Generating beautiful images has never been the exclusive preserve of art: scientific representations of the brain have aesthetic value too, as portrayed in an exhibition at the German Hygiene Museum in Dresden, co-curated with the Moravian Gallery in Brno, Czech Republic. *Images of the Mind* presents more than 200 artistic and scientific works depicting the mind from medieval times to the present day. It also illustrates how the evolving imagery of artists has always been firmly rooted in contemporary scientific knowledge.

Star names among the artists exhibited include Rembrandt, Leonardo da Vinci, Lucas Cranach the Elder and Albrecht Dürer, as well as luminaries Caspar David Friedrich and Edvard Munch, and contemporary artists Bill Viola and Antony Gormley. Unusually, works by relatively unknown artists from central Europe are also on show: Bohumil Kubišta's 1911 *Epileptic Woman* is a masterly portrayal of emotional torment.

No less of a draw is the historical scientific imagery. Sketches by 1906 Nobel prize-winners Camillo Golgi and Santiago Ramón y Cajal are paired to illustrate a legendary scientific dispute. When they examined brain tissue under a microscope, each saw — and drew — a different structure. Golgi sketched a continuous web of cells; Cajal correctly depicted individual cells, or neurons.

Also on view are newly discovered drawings by neuroanatomist Korbinian Brodmann, who mapped the human cerebral cortex in 1908, and stunning images created using Jeff Lichtman's and Joshua Sanes's Brainbow technique. Developed at Harvard University in 2007, the method uses different colours to pick out individual neurons

Images of the Mind
German Hygiene
Museum, Dresden.
Until 30 October.

in a brain slice.

The curators match the quality of the gathered objects with a fascinating narration of the history of mind imagery. The earliest exhibits are eleventh-century manuscripts depicting the Aristotelian understanding of cognitive processes. In the Renaissance, anatomists dissected corpses and drew, beautifully, what they saw, providing templates for artists to paint more realistically. Portraitists quickly attempted to go further, to capture the soul of their subjects as well as their external proportions. Rembrandt's series of tiny, detailed self-portraits in different emotional states encapsulates this perfectly.

Depictions of the mind changed again in the twentieth century, when Sigmund Freud divided the psyche into the conscious and unconscious. Artists such as Kubišta began to explore the fractured mind more abstractly — by then, photography was in any case capturing the realism to which they had formerly striven.

During the past few decades, scientists have described the brain in ever greater electrical, molecular and anatomical detail. Artists have responded, often questioning whether the wet, electrical mass of the physical brain could alone host the mind. Perhaps they are anticipating new neuroscientific understanding of the powerful ways in which environment shapes the brain. For both artists and scientists, the exhibition is a reminder that the mind remains a mysterious moving target. ■

Alison Abbott is Nature's Senior European correspondent.

J. LICHTMAN/HARVARD UNIV.

CORRESPONDENCE

On the care and use of US lab animals

Some recommendations in the latest *Guide for the Care and Use of Laboratory Animals* (see go.nature.com/i9y7fr) have stirred up controversy in the research community. As former members of the committee that updated the guide, we would like to comment informally on these objections (see disclaimer).

The guide was released last year by the US National Research Council. Recommendations may apply to institutions funded by the US Public Health Service, but will, however, be used by the Association for the Assessment and Accreditation of Laboratory Animal Care International, a non-profit organization based in Frederick, Maryland, to evaluate and accredit research institutions.

Because of its many users and settings, the guide is written in broad statements of desired outcomes (performance standards) and is largely devoid of detailed specifications. This enables research institutions to create customized procedures and programmes within the context of the recommendations for high-quality care, from external acquisition or in-house breeding of lab animals to their final disposition.

The latest edition of the guide expands on this approach, partly on the basis of overwhelming support for performance standards by the research community. Despite this, some are urging the National Institutes of Health's Office of Laboratory Animal Welfare (OLAW) not to adopt the eighth edition, spurred by fears over how OLAW will interpret and implement the guide's recommendations. OLAW is currently considering

implementation of the guide and is due to issue a position statement soon.

Objections mainly focus on the guide's space recommendations, particularly for breeding rodents. Optimal space requirements depend on factors including strain, sex, age, enrichment and animal psychology. The guide points out the necessity for study and assessment of the space requirements of laboratory species, while leaving previous recommendations essentially unaltered. Emphasis is placed instead on performance standards that are critical for deciding on minimal cage space.

Lobby groups representing commercial rodent vendors (among others) estimate that compliance with this new standard will cost upwards of US\$500 million. But many institutions have already implemented performance standards to prevent overcrowding or are voluntarily using larger breeding cages. Standards for rodent husbandry should not be bound by commercial profitability.

Further research is needed into space and housing requirements and the provision of enrichment, exercise and human contact. We believe that the latest recommendations and a performance-standards approach will eventually help to define the highest welfare standards for research animals.

Janet C. Garber on behalf of 13 co-signatories*, Garber Consulting, Pinehurst, North Carolina, USA.

jgarber@nc.rr.com

*See <http://dx.doi.org/476152a> for a full list of signatories.

Disclaimer: Opinions expressed in this Correspondence are strictly those of the authors, written in their personal capacity. They do not represent the views of colleagues, current or past employers, or any other body.

Rule on papers puts China's PhDs at risk

To help counter the overemphasis on quantity rather than quality of research publications in China (*Nature* 475, 267; 2011), many Chinese universities require their PhD candidates to publish a specified number of papers in high-impact journals before they are awarded their degree.

Although laudable in principle, this requirement puts even more pressure on students. If that is compounded by inadequate guidance and support from mentors, it will encourage more scientific fraud and hamper the development of junior scientists in the long run. **Derrick Y. F. Lai** McGill University, Montreal, Canada. yukfo.lai@mail.mcgill.ca

Helping hand for genomics in Africa

We offer a personal example of the benefits of lending African scientists 'a helping hand' (*Nature* 474, 542; 2011). After a five-year relationship, a genomics network partnership was last year officially set up between the J. Craig Venter Institute in San Diego, California, and the University of Limpopo, one of the least resourced universities in South Africa.

The University of Limpopo, with campuses in Polokwane and Pretoria, is the result of a merger between South Africa's University of the North and the Medical University of Southern Africa — previously 'non-white' higher-education facilities as defined by the apartheid policies.

The university is striving to build regional resources and a skill base in genomics to help find scientific solutions for rural Africa. International partnerships, such as the one

between our institutions, gives the University of Limpopo's researchers access to current technologies, informatic expertise and training that is unavailable locally.

Together we are seeking practical solutions through educational workshops, student and staff exchanges, improvements in capacity and infrastructure, and equal-contribution projects, while tackling the inevitable challenges as a single unit.

Our experience suggests that 'joining hands' could ultimately prove to be a successful strategy. **Vanessa M. Hayes** J. Craig Venter Institute, San Diego, California, USA. vhayes@jcv.org
Philip A. Venter University of Limpopo, Polokwane, South Africa.
M. Jeffrey Mphahlele University of Limpopo, Pretoria, South Africa.

Mathematics walks into history

Mathematician William Rowan Hamilton did indeed hit on the idea of quaternions — a complex-number system extending into four dimensions — while crossing a Dublin bridge in 1843 (*Nature* 475, 167; 2011). Yet it was what the Irishman was doing while absorbed in mathematics that gives him the last laugh, for quaternions are used in the clinical and laboratory analysis of bodily motion, notably of gait. So what was Hamilton doing? Walking into history.

John Hart Oxford Brookes University, Oxford, UK. johnhart1@btconnect.com

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the author guidelines at go.nature.com/cmchno.

MICROBIOLOGY

Hydrogen for dinner

The vast array of bacterium–animal symbioses at deep-sea hydrothermal vents was thought to be fuelled by just two chemicals. A study of one such symbiosis in its environmental context reveals a third energy source. [SEE ARTICLE P. 176](#)

VICTORIA J. ORPHAN & TORI M. HOEHLER

Considerable efforts have been made in recent years to develop hydrogen as an energy source, with one strategy being to harness the activity of hydrogen-metabolizing microorganisms. On page 176 of this issue, Petersen *et al.*¹ report that humans are not the first animals to use such a strategy: at sea-floor hydrothermal vents, mussels bathed in hydrogen-rich fluids thrive by supporting and harvesting an internal population of hydrogen-consuming bacteria. Hydrogen is therefore the first chemical energy source for microbe–animal symbiosis to have been discovered in 25 years, and joins hydrogen sulphide and methane on the short menu of compounds used to fuel organisms at hydrothermal vents.

Billions of years of photosynthetic activity have left Earth with a thin veneer of highly oxidized material over a mass of chemically reduced rock — which means that the planet is, essentially, a battery. Hydrothermal circulation serves to tap this battery: fluids circulating through the sea-floor crust are charged with reduced (electron-laden) compounds and emerge as jets into the oxidizing environment of the ocean. The transport of compounds from the main reduced rock mass (the anode) to the ocean (the cathode) constitutes a flow of electrons. And, as with any battery, this electron flow is a source of usable energy.

The energy delivered by vent-fluid chemistry fuels thriving communities of microorganisms and animals — densely populated oases of life on an otherwise sparsely inhabited ocean floor (Fig. 1). But some of the vents' most conspicuous and abundant animal inhabitants have neither mouths with which to feed, nor the capability to directly metabolize venting chemicals. Instead, these organisms rely on remarkable symbioses with microorganisms that do have such a capability. The characterization of these unusual bacterium–animal partnerships was initiated 30 years ago, with reports^{2,3} of a symbiosis between the giant tubeworm *Riftia pachyptila* and bacteria that oxidize hydrogen sulphide gas (H₂S). Numerous other examples of symbiotic vent-dwelling animals have since been described. All are based on the metabolism of either methane or hydrogen sulphide⁴.



Figure 1 | In hot water. These mussels and associated fauna are bathed in hydrothermal fluids at the Wideawake vent field on the Mid-Atlantic Ridge. (Photo taken by the remotely operated, deep-sea submersible MARUM-QUEST.)

The seemingly restricted diet of these bacterium–animal symbioses is in many ways surprising. Hydrogen, ammonium, metals and a variety of other reduced compounds can be abundant in vent fluids along with methane and hydrogen sulphide, and free-living microbial consumers of most of these compounds have been identified at vents⁵. Hydrogen, in particular, is a near-ubiquitous product of the reaction of water with oceanic crustal rocks, and is therefore a quantitatively significant component of most vent fluids. In vent systems hosted by peridotite (mantle-like) rock, for example, hydrogen is present in greater abundance, and stores more energy that could potentially be used for microbial metabolism, than either methane or hydrogen sulphide⁶.

The apparent constraints on the chemical basis of bacterium–animal mutualisms could be due to physiological limitations of the host animals⁷. But they could also reflect the fact that our understanding of vent systems is still limited 34 years after their discovery. Petersen *et al.*¹ now reveal that these symbioses are indeed more metabolically versatile than was previously thought.

The study of hydrogen use and other biological and geochemical processes in deep-sea hydrothermal vent systems poses unique challenges because of the remote location of the vents — which, on average, are found at depths greater than 2 kilometres — and the heterogeneous and highly

variable physical and chemical conditions that frequently characterize them. Accessing these environments requires robotic or manned submersibles, but the resulting animal collections, ecological observations and shipboard geochemical analyses often offer only limited 'snapshots' of these dynamic ecosystems. Rather than rely on such discrete shipboard measurements, Petersen *et al.* deployed a mass spectrometer capable of being used *in situ* in the deep sea⁸ to map the distribution of chemicals within vent systems directly. This provided tantalizing evidence that both hydrogen and methane were being consumed within the densely colonized mussel beds.

Although the authors' *in situ* geochemical data were suggestive of active hydrogen metabolism within the mussel beds, further evidence was required to directly link this metabolism to the two known microbial symbionts living in the molluscs' gills. One of these microbes is known to metabolize hydrogen sulphide, whereas the other metabolizes methane (Fig. 2). Petersen and colleagues therefore incubated symbiont-laden gill tissue in sea water infused with hydrogen, and measured the rate at which the microbes fixed carbon dioxide. They found that the rate was similar to the fixation rate of tissues bathed in hydrogen sulphide, indicating that the sulphide-oxidizing and/or methane-oxidizing symbionts were also responsible for the observed hydrogen consumption in the mussel beds.

The authors went on to perform a metagenomic analysis of the symbionts in the gill tissue, and found that a key gene for hydrogen oxidation (*hupL*, which encodes part of the hydrogen-oxidizing enzyme [NiFe]-hydrogenase) is located on a DNA fragment that also contains sulphide-oxidation genes. This strongly suggests that the sulphide-metabolizing symbionts are responsible for hydrogen consumption. To get specific confirmation of this, Petersen *et al.* used fluorescent probes that target the *hupL* gene and the hydrogenase protein in the symbionts. They observed that sulphide-oxidizing symbionts in the gill were lit up by the *hupL* probe, whereas neighbouring methane-oxidizing bacteria were not. Collectively, the authors' observations present a compelling case that the long-known bacterium–mussel symbiosis can operate using a previously unknown metabolism.

It is likely that other symbioses will be discovered that also use hydrogen as a fuel source, given that symbioses involving relatives of the mussels' sulphide-oxidizing symbionts are widespread in reducing environments. Indeed, the authors find¹ that symbionts of the vent-dwelling shrimp *Rimicaris exoculata* have *hupL*, providing genetic evidence that crustaceans might also tap into hydrogen as an energy source through symbiosis. Follow-up studies should now determine the full extent of hydrogen use by symbionts, as well as the environmental and biological factors that influence

when and where hydrogen utilization occurs in chemosynthetic ecosystems.

Petersen and colleagues' work exemplifies the technology-driven revolution that is occurring in the biological sciences. The continuous development of ever more powerful and specific molecular tools allows taxonomic identity and gene content to be linked to metabolic potential and activity, and to be visualized in context. As these techniques converge with new instrumentation that allows the *in situ* characterization of physicochemical parameters — even in environments as remote and extreme as hydrothermal vents — biologists are freed from their reliance on model organisms in artificial surroundings. Now, more than ever, our understanding of biology can be placed in the correct environmental and ecological context, enabling the discovery of previously unknown activities that support life. ■

Victoria J. Orphan is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. **Tori M. Hoehler** is at the NASA Ames Research Center, Moffett Field, California 94035, USA.
e-mails: vorphan@gps.caltech.edu; tori.m.hoehler@nasa.gov

1. Petersen, J. M. *et al.* *Nature* **476**, 176–180 (2011).
2. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. *Science* **213**, 340–342 (1981).
3. Felbeck, H. *Science* **213**, 336–338 (1981).
4. Dubilier, N., Bergin, C. & Lott, C. *Nature Rev. Microbiol.* **6**, 725–740 (2008).
5. Jannasch, H. W. & Mottl, M. *J. Science* **229**, 717–725 (1985).
6. Amend, J. P., McCollom, T. M., Hentscher, M. & Bach, W. *Geochim. Cosmochim. Acta* (in the press).
7. Childress, J. J. & Girguis, P. R. *J. Exp. Biol.* **214**, 312–325 (2011).
8. Wankel, S. D. *et al.* *Nature Geosci.* **4**, 461–468 (2011).

QUANTUM INFORMATION

Microwave ion-trap quantum computing

A new type of ion-trap quantum technology has been developed that uses microwave radiation to perform computations. It will considerably simplify the practical implementation of large-scale quantum computers. SEE LETTERS P.181 & P.185

WINFRIED K. HENSINGER

The strange phenomena of quantum physics, such as the possibility of a single atom being in two different places simultaneously, have mystified experienced physicists and students alike. These phenomena are not just theoretical curiosities: a new, practical, field of quantum physics has become particularly vibrant during the past ten years. It concerns harnessing the power of quantum effects to produce an innovative type of technology — quantum technology. This has the potential to revolutionize computing. Although ways to implement this technology are being pursued in various physical systems, quantum computing using trapped ions^{1–3} has undoubtedly been the most successful so far. In this issue, Ospelkaus *et al.*⁴ (page 181) and Timoney *et al.*⁵ (page 185) describe a trapped-ion approach that further bolsters the enormous potential of this system for the implementation of large-scale quantum computers.

Ospelkaus *et al.*⁴ report the realization of quantum gates using microwave radiation instead of laser beams, which have until now been used to implement such gates. Quantum gates are the analogues of classical logic gates, and are used to perform computations in a quantum processor. Timoney *et al.*⁵

discuss a new approach to quantum computing — involving microwaves — that is highly resilient to noise (often referred to as decoherence) acting on the physical system used to execute the computations. Decoherence has the potential to destroy quantum effects and interfere with the operation of a quantum processor. Timoney and colleagues' scheme effectively shields the quantum processor from decoherence.

Quantum gates executed in a quantum processor are based on the creation of entanglement. Entanglement is one of the non-intuitive phenomena of quantum physics, whereby the properties of multiple systems, such as groups of ions, are correlated. Ordinarily, the outcome of measuring a particular property on an ion yields a completely random result. But measuring the same property on each ion of a set of entangled ions produces correlated results. Non-entangled ions do not produce such correlations. An easy way to understand the situation is to imagine two people each tossing a coin. The outcome of each coin toss is random. However, if the coins were entangled in a particular way, the outcome of the coin toss would always be the same for both coins — with both people getting heads or tails.

The past few years have witnessed ground-

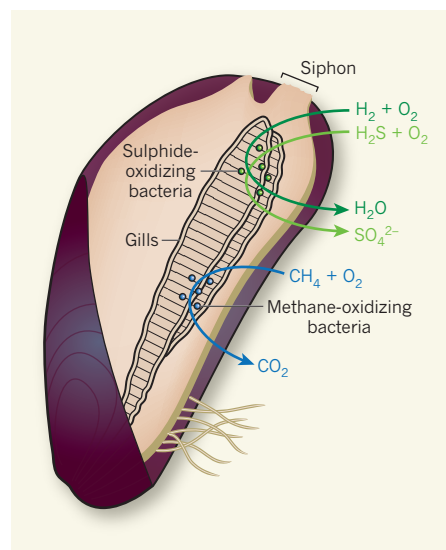


Figure 2 | Hydrogen-consuming symbiosis.

Bathymodiolus puteoserpentis mussels live at hydrothermal deep-sea vents, which emit hydrogen, methane (CH_4) and hydrogen sulphide (H_2S), among other chemicals. The mussels draw water through their siphons and into enlarged respiratory gills that house symbiotic bacteria. Some of these bacteria harness energy by oxidizing methane to carbon dioxide (blue arrow), whereas others do this by oxidizing hydrogen sulphide to sulphate (SO_4^{2-} , light green arrow). Petersen *et al.*¹ report that the sulphide-oxidizing bacteria also use hydrogen as an energy source, oxidizing it to water (dark green arrow).

The authors went on to perform a metagenomic analysis of the symbionts in the gill tissue, and found that a key gene for hydrogen oxidation (*hupL*, which encodes part of the hydrogen-oxidizing enzyme [NiFe]-hydrogenase) is located on a DNA fragment that also contains sulphide-oxidation genes. This strongly suggests that the sulphide-metabolizing symbionts are responsible for hydrogen consumption. To get specific confirmation of this, Petersen *et al.* used fluorescent probes that target the *hupL* gene and the hydrogenase protein in the symbionts. They observed that sulphide-oxidizing symbionts in the gill were lit up by the *hupL* probe, whereas neighbouring methane-oxidizing bacteria were not. Collectively, the authors' observations present a compelling case that the long-known bacterium–mussel symbiosis can operate using a previously unknown metabolism.

It is likely that other symbioses will be discovered that also use hydrogen as a fuel source, given that symbioses involving relatives of the mussels' sulphide-oxidizing symbionts are widespread in reducing environments. Indeed, the authors find¹ that symbionts of the vent-dwelling shrimp *Rimicaris exoculata* have *hupL*, providing genetic evidence that crustaceans might also tap into hydrogen as an energy source through symbiosis. Follow-up studies should now determine the full extent of hydrogen use by symbionts, as well as the environmental and biological factors that influence

when and where hydrogen utilization occurs in chemosynthetic ecosystems.

Petersen and colleagues' work exemplifies the technology-driven revolution that is occurring in the biological sciences. The continuous development of ever more powerful and specific molecular tools allows taxonomic identity and gene content to be linked to metabolic potential and activity, and to be visualized in context. As these techniques converge with new instrumentation that allows the *in situ* characterization of physicochemical parameters — even in environments as remote and extreme as hydrothermal vents — biologists are freed from their reliance on model organisms in artificial surroundings. Now, more than ever, our understanding of biology can be placed in the correct environmental and ecological context, enabling the discovery of previously unknown activities that support life. ■

Victoria J. Orphan is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. **Tori M. Hoehler** is at the NASA Ames Research Center, Moffett Field, California 94035, USA.
e-mails: vorphan@gps.caltech.edu; tori.m.hoehler@nasa.gov

1. Petersen, J. M. *et al.* *Nature* **476**, 176–180 (2011).
2. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. *Science* **213**, 340–342 (1981).
3. Felbeck, H. *Science* **213**, 336–338 (1981).
4. Dubilier, N., Bergin, C. & Lott, C. *Nature Rev. Microbiol.* **6**, 725–740 (2008).
5. Jannasch, H. W. & Mottl, M. *J. Science* **229**, 717–725 (1985).
6. Amend, J. P., McCollom, T. M., Hentscher, M. & Bach, W. *Geochim. Cosmochim. Acta* (in the press).
7. Childress, J. J. & Girguis, P. R. *J. Exp. Biol.* **214**, 312–325 (2011).
8. Wankel, S. D. *et al.* *Nature Geosci.* **4**, 461–468 (2011).

QUANTUM INFORMATION

Microwave ion-trap quantum computing

A new type of ion-trap quantum technology has been developed that uses microwave radiation to perform computations. It will considerably simplify the practical implementation of large-scale quantum computers. SEE LETTERS P.181 & P.185

WINFRIED K. HENSINGER

The strange phenomena of quantum physics, such as the possibility of a single atom being in two different places simultaneously, have mystified experienced physicists and students alike. These phenomena are not just theoretical curiosities: a new, practical, field of quantum physics has become particularly vibrant during the past ten years. It concerns harnessing the power of quantum effects to produce an innovative type of technology — quantum technology. This has the potential to revolutionize computing. Although ways to implement this technology are being pursued in various physical systems, quantum computing using trapped ions^{1–3} has undoubtedly been the most successful so far. In this issue, Ospelkaus *et al.*⁴ (page 181) and Timoney *et al.*⁵ (page 185) describe a trapped-ion approach that further bolsters the enormous potential of this system for the implementation of large-scale quantum computers.

Ospelkaus *et al.*⁴ report the realization of quantum gates using microwave radiation instead of laser beams, which have until now been used to implement such gates. Quantum gates are the analogues of classical logic gates, and are used to perform computations in a quantum processor. Timoney *et al.*⁵

discuss a new approach to quantum computing — involving microwaves — that is highly resilient to noise (often referred to as decoherence) acting on the physical system used to execute the computations. Decoherence has the potential to destroy quantum effects and interfere with the operation of a quantum processor. Timoney and colleagues' scheme effectively shields the quantum processor from decoherence.

Quantum gates executed in a quantum processor are based on the creation of entanglement. Entanglement is one of the non-intuitive phenomena of quantum physics, whereby the properties of multiple systems, such as groups of ions, are correlated. Ordinarily, the outcome of measuring a particular property on an ion yields a completely random result. But measuring the same property on each ion of a set of entangled ions produces correlated results. Non-entangled ions do not produce such correlations. An easy way to understand the situation is to imagine two people each tossing a coin. The outcome of each coin toss is random. However, if the coins were entangled in a particular way, the outcome of the coin toss would always be the same for both coins — with both people getting heads or tails.

The past few years have witnessed ground-

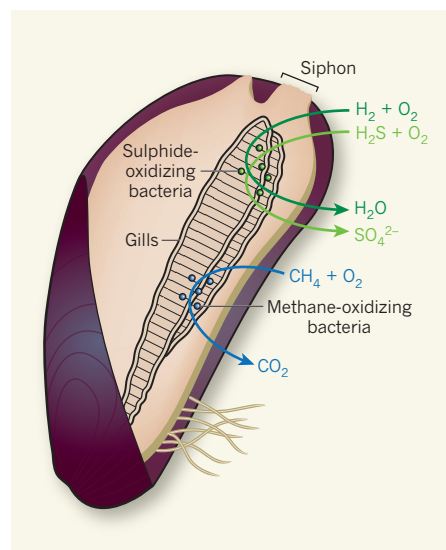


Figure 2 | Hydrogen-consuming symbiosis.

Bathymodiolus puteoserpentis mussels live at hydrothermal deep-sea vents, which emit hydrogen, methane (CH_4) and hydrogen sulphide (H_2S), among other chemicals. The mussels draw water through their siphons and into enlarged respiratory gills that house symbiotic bacteria. Some of these bacteria harness energy by oxidizing methane to carbon dioxide (blue arrow), whereas others do this by oxidizing hydrogen sulphide to sulphate (SO_4^{2-} , light green arrow). Petersen *et al.*¹ report that the sulphide-oxidizing bacteria also use hydrogen as an energy source, oxidizing it to water (dark green arrow).



50 Years Ago

The Prehistoric Chamber Tombs of France. By Glyn Daniel — To win appreciation for some phase of antiquity on the strength of modern excavations is far easier than what has been undertaken here — where the antiquities concerned have been known for centuries, until only yesterday often excavated badly, and celebrated in a literature in which perverse and obsolete terminologies have run riot. Dr. Daniel, supported by his wife, and with backing from many quarters as well as friendly French co-operation, has for long been working towards a systematic account of the megalithic and related stone tombs of France, their form and contents, and their placing in the frame of European prehistory in the third and second millennia B.C., to which in general they belong. This book, by no means his first study of megalithic structures, is his best so far. It is not too hard to read.

From *Nature* 12 August 1961

100 Years Ago

In the July number of *The American Naturalist* Dr. O. P. Hay reopens the discussion with regard to the position of the limbs in *Diplodocus* and other sauropod dinosaurs, criticising the views of those who assert that these reptiles carried themselves in elephantine fashion, and maintaining his own opinion that the general pose was more after the crocodilian style ... Mr. Hay expresses doubts as to whether the erect bird-like posture attributed to the carnivorous dinosaurs of the Jurassic is really true to nature ... In reference to the opinion of Dr. Matthew that Sauropods were too bulky to have lived on land, it is added that "the law to which he gives expression does, of course, prescribe a limit to the size an animal can attain, but who has yet determined what that limit is?"

From *Nature* 10 August 1911

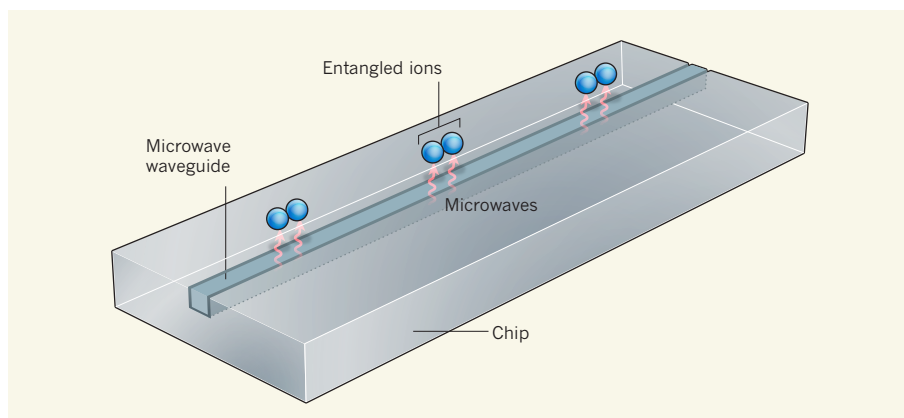


Figure 1 | Microwave ion-trap chip. Ospelkaus and colleagues' approach⁴ to entangling pairs of ions and generating a quantum gate involves launching microwaves into a waveguide incorporated on a chip.

breaking advances in quantum-information processing using trapped ions, including the entanglement of up to 14 ions⁶, and the development of other entanglement-based protocols, such as the realization of a number of quantum algorithms³ and of teleportation^{7,8}. These advances have been made in experiments in which laser beams are used to perform entanglement operations. In 2001, Mintert and Wunderlich had the visionary idea⁹ of implementing quantum gates using long-wavelength radiation, such as microwaves or radio waves. Whereas laser beams must be carefully aligned to interact with the trapped ions that are to be entangled, microwaves can be applied via waveguides (structures that can guide radiation) that are part of the chip on which the ion trap is integrated¹⁰, and so do not require alignment.

What's more, it is much easier and less costly to generate microwave radiation than it is to use the complicated laser systems currently employed, and highly stable microwave sources are readily available. Large-scale quantum computers may require many millions of individually trapped ions, each constituting a single quantum bit (the basic unit of information storage in a quantum computer). As a result, creating the required number of laser beams to entangle the ions may entail significant engineering and come at a considerable cost. By contrast, the use of microwave radiation for the same purpose would be much easier and would make the construction of a large-scale ion-trap quantum-information processor much simpler.

Based on their proposal¹¹ to make use of the oscillating magnetic fields that are inherent to microwave radiation (the original proposal by Mintert and Wunderlich⁹ requires static magnetic fields in addition to microwaves), Ospelkaus *et al.*⁴ have realized the first microwave quantum gate. They achieved this by using a waveguide integrated into a microchip (Fig. 1) that holds the ion-trap structure. The microchip contains electrodes that produce electric fields capable of trapping two ions just above the chip's surface. Multiple pulses of microwave radiation are then applied to the trapped

ions through the waveguide, effectively entangling the two ions and successfully executing a quantum gate.

Meanwhile, Timoney *et al.*⁵ trapped individual ions and applied a number of microwave pulses to them. This approach sets the ions to a state in which they are decoupled from outside noise. An easy way to visualize this is by considering the suspension of a common car. Springs in the car's suspension system decouple the car frame from the wheels, largely isolating the driver from vibrations caused by uneven road surfaces. In a similar way, Timoney and colleagues' scheme allows trapped ions to be isolated from external disturbances that would otherwise have the potential to disturb the operation of a microwave trapped-ion quantum processor.

The achievements of Ospelkaus *et al.*⁴ and Timoney *et al.*⁵ constitute step-changing innovations for quantum computing with trapped ions because they will probably aid the production of large-scale ion-trap quantum computers on foreseeable timescales. Quantum computing is likely to revolutionize many areas of science, and we have only just started to appreciate its true potential. ■

Winfried K. Hensinger is in the Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK.
e-mail: w.k.hensinger@sussex.ac.uk

1. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Wineland, D. J. *et al.* *J. Res. Natl. Inst. Stand. Technol.* **103**, 259–328 (1998).
3. Häffner, H., Roos, C. F. & Blatt, R. *Phys. Rep.* **469**, 155–203 (2008).
4. Ospelkaus, C. *et al.* *Nature* **476**, 181–184 (2011).
5. Timoney, N. *et al.* *Nature* **476**, 185–188 (2011).
6. Monz, T. *et al.* *Phys. Rev. Lett.* **106**, 130506 (2011).
7. Leibfried, D. *et al.* *Nature* **422**, 412–415 (2003).
8. Schmidt-Kaler, F. *et al.* *Nature* **422**, 408–411 (2003).
9. Mintert, F. & Wunderlich, C. *Phys. Rev. Lett.* **87**, 257904 (2001).
10. Hughes, M. D., Lekitsch, B., Broersma, J. A. & Hensinger, W. K. *Contemp. Phys.* <http://dx.doi.org/10.1080/00107514.2011.601918>; preprint available at arXiv:1101.3207v2 [quant-ph] (2011).
11. Ospelkaus, C. *et al.* *Phys. Rev. Lett.* **101**, 090502 (2008).



50 Years Ago

The Prehistoric Chamber Tombs of France. By Glyn Daniel — To win appreciation for some phase of antiquity on the strength of modern excavations is far easier than what has been undertaken here — where the antiquities concerned have been known for centuries, until only yesterday often excavated badly, and celebrated in a literature in which perverse and obsolete terminologies have run riot. Dr. Daniel, supported by his wife, and with backing from many quarters as well as friendly French co-operation, has for long been working towards a systematic account of the megalithic and related stone tombs of France, their form and contents, and their placing in the frame of European prehistory in the third and second millennia B.C., to which in general they belong. This book, by no means his first study of megalithic structures, is his best so far. It is not too hard to read.

From *Nature* 12 August 1961

100 Years Ago

In the July number of *The American Naturalist* Dr. O. P. Hay reopens the discussion with regard to the position of the limbs in *Diplodocus* and other sauropod dinosaurs, criticising the views of those who assert that these reptiles carried themselves in elephantine fashion, and maintaining his own opinion that the general pose was more after the crocodilian style ... Mr. Hay expresses doubts as to whether the erect bird-like posture attributed to the carnivorous dinosaurs of the Jurassic is really true to nature ... In reference to the opinion of Dr. Matthew that Sauropods were too bulky to have lived on land, it is added that "the law to which he gives expression does, of course, prescribe a limit to the size an animal can attain, but who has yet determined what that limit is?"

From *Nature* 10 August 1911

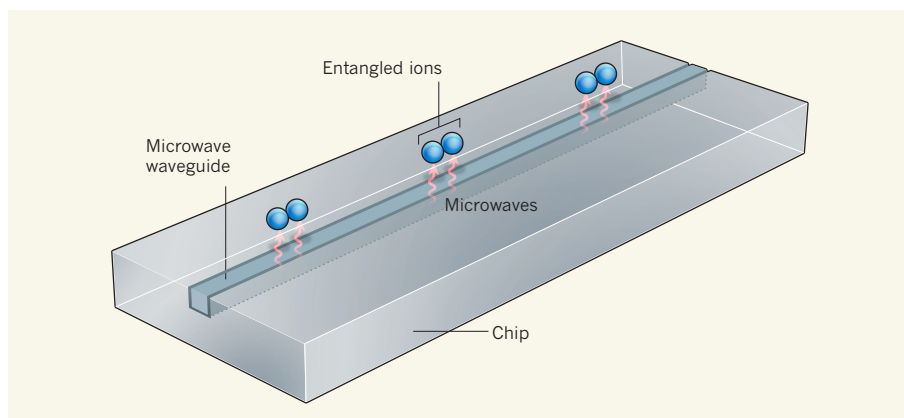


Figure 1 | Microwave ion-trap chip. Ospelkaus and colleagues' approach⁴ to entangling pairs of ions and generating a quantum gate involves launching microwaves into a waveguide incorporated on a chip.

breaking advances in quantum-information processing using trapped ions, including the entanglement of up to 14 ions⁶, and the development of other entanglement-based protocols, such as the realization of a number of quantum algorithms³ and of teleportation^{7,8}. These advances have been made in experiments in which laser beams are used to perform entanglement operations. In 2001, Mintert and Wunderlich had the visionary idea⁹ of implementing quantum gates using long-wavelength radiation, such as microwaves or radio waves. Whereas laser beams must be carefully aligned to interact with the trapped ions that are to be entangled, microwaves can be applied via waveguides (structures that can guide radiation) that are part of the chip on which the ion trap is integrated¹⁰, and so do not require alignment.

What's more, it is much easier and less costly to generate microwave radiation than it is to use the complicated laser systems currently employed, and highly stable microwave sources are readily available. Large-scale quantum computers may require many millions of individually trapped ions, each constituting a single quantum bit (the basic unit of information storage in a quantum computer). As a result, creating the required number of laser beams to entangle the ions may entail significant engineering and come at a considerable cost. By contrast, the use of microwave radiation for the same purpose would be much easier and would make the construction of a large-scale ion-trap quantum-information processor much simpler.

Based on their proposal¹¹ to make use of the oscillating magnetic fields that are inherent to microwave radiation (the original proposal by Mintert and Wunderlich⁹ requires static magnetic fields in addition to microwaves), Ospelkaus *et al.*⁴ have realized the first microwave quantum gate. They achieved this by using a waveguide integrated into a microchip (Fig. 1) that holds the ion-trap structure. The microchip contains electrodes that produce electric fields capable of trapping two ions just above the chip's surface. Multiple pulses of microwave radiation are then applied to the trapped

ions through the waveguide, effectively entangling the two ions and successfully executing a quantum gate.

Meanwhile, Timoney *et al.*⁵ trapped individual ions and applied a number of microwave pulses to them. This approach sets the ions to a state in which they are decoupled from outside noise. An easy way to visualize this is by considering the suspension of a common car. Springs in the car's suspension system decouple the car frame from the wheels, largely isolating the driver from vibrations caused by uneven road surfaces. In a similar way, Timoney and colleagues' scheme allows trapped ions to be isolated from external disturbances that would otherwise have the potential to disturb the operation of a microwave trapped-ion quantum processor.

The achievements of Ospelkaus *et al.*⁴ and Timoney *et al.*⁵ constitute step-changing innovations for quantum computing with trapped ions because they will probably aid the production of large-scale ion-trap quantum computers on foreseeable timescales. Quantum computing is likely to revolutionize many areas of science, and we have only just started to appreciate its true potential. ■

Winfried K. Hensinger is in the Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK.
e-mail: w.k.hensinger@sussex.ac.uk

1. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Wineland, D. J. *et al.* *J. Res. Natl. Inst. Stand. Technol.* **103**, 259–328 (1998).
3. Häffner, H., Roos, C. F. & Blatt, R. *Phys. Rep.* **469**, 155–203 (2008).
4. Ospelkaus, C. *et al.* *Nature* **476**, 181–184 (2011).
5. Timoney, N. *et al.* *Nature* **476**, 185–188 (2011).
6. Monz, T. *et al.* *Phys. Rev. Lett.* **106**, 130506 (2011).
7. Leibfried, D. *et al.* *Nature* **422**, 412–415 (2003).
8. Schmidt-Kaler, F. *et al.* *Nature* **422**, 408–411 (2003).
9. Mintert, F. & Wunderlich, C. *Phys. Rev. Lett.* **87**, 257904 (2001).
10. Hughes, M. D., Lekitsch, B., Broersma, J. A. & Hensinger, W. K. *Contemp. Phys.* <http://dx.doi.org/10.1080/00107514.2011.601918>; preprint available at arXiv:1101.3207v2 [quant-ph] (2011).
11. Ospelkaus, C. *et al.* *Phys. Rev. Lett.* **101**, 090502 (2008).

ATMOSPHERIC SCIENCE

Enigma of the recent methane budget

The previously increasing atmospheric methane concentration has inexplicably stalled over the past three decades. This may be due to a fall in fossil-fuel emissions or to farming practices that are curtailing microbial sources. [SEE LETTERS P.194 & P.198](#)

MARTIN HEIMANN

There has been a recent unexpected variation in the atmospheric concentration of methane (CH_4) — the second most important anthropogenically modified greenhouse gas after carbon dioxide. Its growth rate has declined almost to a standstill in the past three decades, despite more than doubling since pre-industrial times; signs of renewed growth are only now starting to appear. What has caused the puzzling behaviour of this atmospheric trace gas? In potentially contradictory studies on pages 194 and 198 of this issue, Kai *et al.*¹ and Aydin *et al.*² attempt to answer this question.

The growth rate of the atmospheric methane concentration effectively reflects the global methane budget (Fig. 1), which at present delicately balances large sinks and sources. Global sinks arise primarily from the activity of the hydroxyl radical (OH^\bullet), which is involved in photochemical oxidation reactions in the atmosphere. Although sinks are quite well understood, the relative contributions of the various processes that produce methane are uncertain.

About 60% of global methane stems from human activities. This includes emissions associated with energy production and usage — such as coalmining, incomplete combustion of fossil fuels and gas leaks. These fossil-fuel-derived emissions, together with methane from landfills, organic waste, cattle raising, rice agriculture and biomass burning, were probably the main cause of the increasing concentrations after pre-industrial times. Methane also forms naturally, mostly from anaerobic decomposition of organic carbon in wetlands, with lesser contributions from the ocean, termites, wild animals, wildfires and various geological sources.

How can the relative contributions of the various sources to the global methane budget be determined? One technique depends on the fact that most of the source processes have a distinct geographical signature, which is reflected in the atmospheric methane-concentration distribution. For example, fossil-fuel methane emissions that predominantly occur in the Northern Hemisphere cause a north-south concentration gradient that is visible in

atmospheric measurements, including those made at surface stations³ and from space-based observations⁴. This method identified a fall in fossil-fuel-derived methane emissions in the 1990s, reinforced by a coincident decrease in wetland emissions since 2000, as the main drivers of the global decline in methane growth rate over the past 30 years⁵, albeit with substantial uncertainties. Other methane sources also have distinct geographical patterns, but many of these overlap, and so the attribution and quantification of the different sources from atmospheric measurements are not unique.

Aydin and colleagues² offer new information on the history of methane emissions from fossil fuels in the Northern and Southern Hemispheres over the past 60 years. They achieve their reconstruction by measuring the ethane (C_2H_6) in air trapped in porous snow-ice (firn) in Greenland and Antarctica. Like methane, ethane is released during the mining, transport and incomplete burning of fossil fuels, but also from incomplete biomass burning. With the help of a simple atmospheric mixing model, the authors resolve the ethane-concentration

record into changes in ethane-source strength and, given typical hemispheric patterns of fossil-fuel and biomass-burning sources, into changes over time of these two sources. They then assume fixed methane-to-ethane emission ratios in estimating the history of fossil-fuel methane emissions.

Their results are surprising. Whereas the emissions deduced for biomass burning are consistent with independent bottom-up estimates, the inferred history of fossil-fuel-derived methane emissions before 1980 is strikingly different — double the estimates from standard databases based on the statistics of fossil-fuel production. During 1980–2000, the record of fossil-fuel methane emissions shows an almost 30% decline, which would go a long way towards explaining the observed decrease in the global methane growth rate.

Kai and colleagues¹ offer an alternative explanation. They make use of the fact that methane from fossil fuels is enriched in its stable $^{13}\text{C}/^{12}\text{C}$ carbon-isotope ratio, whereas methane from microbial sources (mainly wetlands and rice paddies) is depleted in $^{13}\text{C}/^{12}\text{C}$ with respect to the atmospheric background. Furthermore, variations in the $^2\text{H}/^1\text{H}$ hydrogen-isotope ratio in methane are primarily affected by changes in the photochemical sink. Observations of the atmospheric methane concentration, in conjunction with measurements of its stable-isotope ratios, thus provide an alternative means to distinguish between different source categories. In combination with a simple model of atmospheric mixing, the authors compare the simulated atmospheric signatures of various methane-source scenarios with observations from the past two decades in each hemisphere.

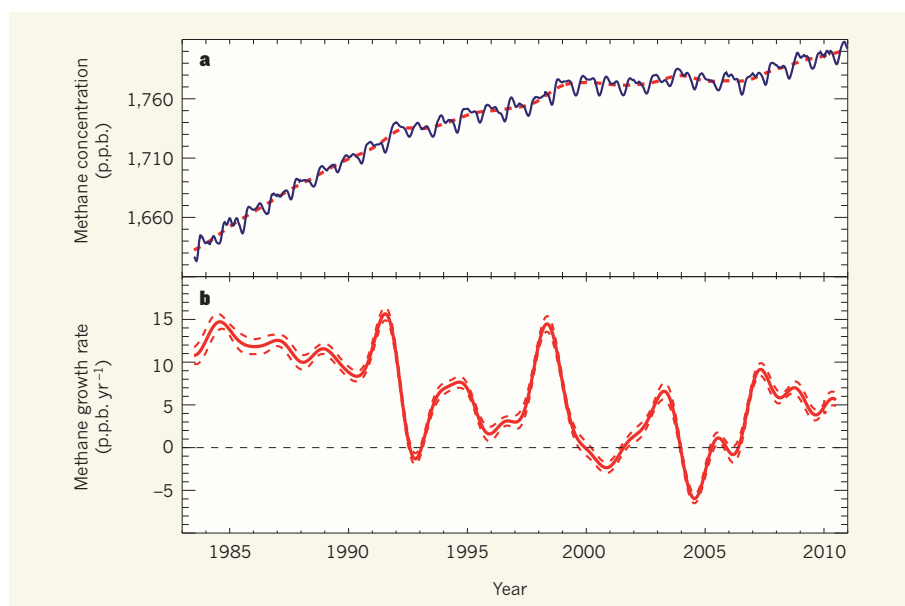


Figure 1 | Changes in global atmospheric methane over the past three decades. **a**, Globally averaged methane concentration, compiled from a global network of atmospheric measurement stations. **b**, Globally averaged growth rate of methane in the atmosphere. Dashed lines indicate ± 1 standard deviation; p.p.b., parts per billion. (Data and graphic are from E. Dlugokencky³, NOAA-GMD.)

Surprisingly, Kai *et al.* find that a reduction in the fossil-fuel methane source is not compatible with these measurements, and that the isotope record can be explained only by a reduction in the microbial sources in the Northern Hemisphere. A drying trend in northern wetlands could have contributed to this finding⁶, but the authors convincingly show that methane emissions from rice agriculture, particularly in China, must also have decreased. Their conclusion is based on changes in agricultural practices: new high-yield rice species, together with greater application of fertilizer, require shorter inundation periods, making substantial water savings and reducing methane emissions.

Can these conflicting inferences on the recent slow-down of global methane growth be

reconciled? Because of the limited data coverage and simplistic analysis assumptions of the two studies^{1,2}, there are considerable uncertainties in the deduced methane-source variations, but the different scenarios are plausible and compatible with their respective observations. The challenge now is to bring the different lines of evidence together, perhaps by using a more advanced modelling framework and improved bottom-up inventory information on the various methane-source categories. More extended observations will help too — particularly the mapping of atmospheric methane concentration by current and upcoming satellite missions.

These studies^{1,2} illustrate the importance of high-precision, long-term observations of methane concentration and isotope composition, and of auxiliary trace gases such

as ethane, in distinguishing between the contributions from different sources. But more insight is needed to solve the enigma of the recent methane budget if the evolution of this important greenhouse gas over the twenty-first century is to be predicted. ■

Martin Heimann is at the Max Planck Institute for Biogeochemistry, 07701 Jena, Germany.
e-mail: martin.heimann@bgc-jena.mpg.de

1. Kai, F. M., Tyler, S. C., Randerson, J. T. & Blake, D. R. *Nature* **476**, 194–197 (2011).
2. Aydin, M. *et al.* *Nature* **476**, 198–201 (2011).
3. Dlugokencky, E. J. *et al.* *Geophys. Res. Lett.* **36**, L18803; <http://dx.doi.org/10.1029/2009GL039780> (2009).
4. Schneising, O. *et al.* *Atmos. Chem. Phys.* **9**, 443–465 (2009).
5. Bousquet, P. *et al.* *Nature* **443**, 439–443 (2006).
6. Jung, M. *et al.* *Nature* **467**, 951–954 (2010).

another team⁶ that recently reported the direct conversion of human fibroblasts into dopaminergic cells using a different combination of transcription factors.

Yoo *et al.*³ (page 228) approached the search for a suitable conversion method from a different angle. They show that combined expression of miR-9/9* and miR-124 — two members of a class of short regulatory RNA sequences called microRNAs (miRNAs) — is sufficient to convert human fibroblasts into neurons. Again, the conversion process was greatly enhanced by the introduction of another bHLH-family transcription factor, NeuroD2. And increased expression of Ascl1 and Myt1l further increased the conversion efficiency of neonatal human fibroblasts: about 80% of the resulting neurons showed action-potential-like activity. As in the other two studies, Yoo and colleagues write that adult fibroblasts were less amenable to conversion.

Several common themes emerge. The studies^{1–3} highlight the importance of the bHLH-family transcription factors and the role of miRNAs in the conversion process. Moreover, they all show that the conversion of mouse and human fibroblasts into neurons requires different protocols, and that conversion of fibroblasts from adult or aged individuals is much less productive than that of embryonic or neonatal fibroblasts.

The essential role of NeuroD1 and NeuroD2 for efficient conversion correlates well with the crucial part played by these two transcription factors in neuron differentiation during normal brain development, and their reduced expression as neurogenesis decreases during ageing^{7,8}. A relevant question is whether the degree of methylation of the genomic region containing the genes for these transcription factors is higher in human than in mouse fibroblasts and increases during ageing, thereby resulting in their reduced expression. DNA methylation is a crucial regulator of gene expression.

How miR-9/9* and miR-124 substitute for Brn2 — and possibly for Ascl1 and Myt1l — in

REGENERATIVE MEDICINE

Bespoke cells for the human brain

Human skin cells have been directly converted into neurons, an achievement that could lead to the cell-based treatment of neurodegenerative disorders. But the road ahead remains long and tortuous. SEE LETTERS P.220, P.224 & P.228

MICHAEL SENDTNER

A major goal for biomedical researchers has been to repair damaged cells and tissues by reinitiating the body's developmental mechanisms. In this respect, the use of embryonic stem cells — and of induced pluripotent stem cells, which are generated by reprogramming differentiated adult cells — has seemed promising. But these cells come with various problems, including ethical concerns, and potential tumorigenicity and rejection by the host immune system. Three papers in this issue^{1–3} describe the direct conversion of neurons from skin fibroblast cells. There is hope, although no compelling evidence, that neurons generated in this way might be superior to those generated from induced pluripotent stem cells, thereby sidestepping the problems of using such cells⁴.

The conversion of mouse fibroblasts into neurons was reported last year⁵. The same team (Pang *et al.*¹, page 220) now find, however, that introducing the same three transcription factors (Brn2, Ascl1/Mash1 and Myt1l) used in their original study into human fibroblasts is not sufficient to convert the cells into functional neurons; the resulting cells are immature, and even after 3 weeks in culture do not become functional, as judged by measurements of action potentials. Nonetheless, the authors show that the efficiency of human

fibroblast conversion into neurons is enhanced by including an extra transcription factor, NeuroD1, a member of the bHLH family.

Caiazzo *et al.*² (page 224) also set out to convert human fibroblasts into a specific type of neuron, in this case neurons of the midbrain that secrete the neurotransmitter dopamine. Starting their study in mice, they tested the efficiency of mixtures of transcription factors (combinations of Ascl1, Brn2, Myt1l and others) in the conversion process. The winning combination was that of Ascl1, Nurr1 and Lmx1a; Brn2 and Myt1l did not lead to dopamine-secreting neurons. The authors provide evidence that the conversion route was direct, and that the resulting neurons were functionally surprisingly similar to primary dopaminergic neurons isolated from the brain. What's more, when the converted dopaminergic cells were grafted into neonatal mouse brains, they became integrated and were functional.

Caiazzo and colleagues then applied their method to human fibroblasts, including those from patients with Parkinson's disease, a neurodegenerative disorder in which dopaminergic neurons are damaged. The conversion efficiency of adult human fibroblasts was 10–20-fold lower than that of their mouse equivalents, and the resulting cells were less mature. This discrepancy points to species differences between mice and humans². The authors' observations fit well with those of

Surprisingly, Kai *et al.* find that a reduction in the fossil-fuel methane source is not compatible with these measurements, and that the isotope record can be explained only by a reduction in the microbial sources in the Northern Hemisphere. A drying trend in northern wetlands could have contributed to this finding⁶, but the authors convincingly show that methane emissions from rice agriculture, particularly in China, must also have decreased. Their conclusion is based on changes in agricultural practices: new high-yield rice species, together with greater application of fertilizer, require shorter inundation periods, making substantial water savings and reducing methane emissions.

Can these conflicting inferences on the recent slow-down of global methane growth be

reconciled? Because of the limited data coverage and simplistic analysis assumptions of the two studies^{1,2}, there are considerable uncertainties in the deduced methane-source variations, but the different scenarios are plausible and compatible with their respective observations. The challenge now is to bring the different lines of evidence together, perhaps by using a more advanced modelling framework and improved bottom-up inventory information on the various methane-source categories. More extended observations will help too — particularly the mapping of atmospheric methane concentration by current and upcoming satellite missions.

These studies^{1,2} illustrate the importance of high-precision, long-term observations of methane concentration and isotope composition, and of auxiliary trace gases such

as ethane, in distinguishing between the contributions from different sources. But more insight is needed to solve the enigma of the recent methane budget if the evolution of this important greenhouse gas over the twenty-first century is to be predicted. ■

Martin Heimann is at the Max Planck Institute for Biogeochemistry, 07701 Jena, Germany.
e-mail: martin.heimann@bgc-jena.mpg.de

1. Kai, F. M., Tyler, S. C., Randerson, J. T. & Blake, D. R. *Nature* **476**, 194–197 (2011).
2. Aydin, M. *et al.* *Nature* **476**, 198–201 (2011).
3. Dlugokencky, E. J. *et al.* *Geophys. Res. Lett.* **36**, L18803; <http://dx.doi.org/10.1029/2009GL039780> (2009).
4. Schneising, O. *et al.* *Atmos. Chem. Phys.* **9**, 443–465 (2009).
5. Bousquet, P. *et al.* *Nature* **443**, 439–443 (2006).
6. Jung, M. *et al.* *Nature* **467**, 951–954 (2010).

another team⁶ that recently reported the direct conversion of human fibroblasts into dopaminergic cells using a different combination of transcription factors.

Yoo *et al.*³ (page 228) approached the search for a suitable conversion method from a different angle. They show that combined expression of miR-9/9* and miR-124 — two members of a class of short regulatory RNA sequences called microRNAs (miRNAs) — is sufficient to convert human fibroblasts into neurons. Again, the conversion process was greatly enhanced by the introduction of another bHLH-family transcription factor, NeuroD2. And increased expression of Ascl1 and Myt1l further increased the conversion efficiency of neonatal human fibroblasts: about 80% of the resulting neurons showed action-potential-like activity. As in the other two studies, Yoo and colleagues write that adult fibroblasts were less amenable to conversion.

Several common themes emerge. The studies^{1–3} highlight the importance of the bHLH-family transcription factors and the role of miRNAs in the conversion process. Moreover, they all show that the conversion of mouse and human fibroblasts into neurons requires different protocols, and that conversion of fibroblasts from adult or aged individuals is much less productive than that of embryonic or neonatal fibroblasts.

The essential role of NeuroD1 and NeuroD2 for efficient conversion correlates well with the crucial part played by these two transcription factors in neuron differentiation during normal brain development, and their reduced expression as neurogenesis decreases during ageing^{7,8}. A relevant question is whether the degree of methylation of the genomic region containing the genes for these transcription factors is higher in human than in mouse fibroblasts and increases during ageing, thereby resulting in their reduced expression. DNA methylation is a crucial regulator of gene expression.

How miR-9/9* and miR-124 substitute for Brn2 — and possibly for Ascl1 and Myt1l — in

REGENERATIVE MEDICINE

Bespoke cells for the human brain

Human skin cells have been directly converted into neurons, an achievement that could lead to the cell-based treatment of neurodegenerative disorders. But the road ahead remains long and tortuous. SEE LETTERS P.220, P.224 & P.228

MICHAEL SENDTNER

A major goal for biomedical researchers has been to repair damaged cells and tissues by reinitiating the body's developmental mechanisms. In this respect, the use of embryonic stem cells — and of induced pluripotent stem cells, which are generated by reprogramming differentiated adult cells — has seemed promising. But these cells come with various problems, including ethical concerns, and potential tumorigenicity and rejection by the host immune system. Three papers in this issue^{1–3} describe the direct conversion of neurons from skin fibroblast cells. There is hope, although no compelling evidence, that neurons generated in this way might be superior to those generated from induced pluripotent stem cells, thereby sidestepping the problems of using such cells⁴.

The conversion of mouse fibroblasts into neurons was reported last year⁵. The same team (Pang *et al.*¹, page 220) now find, however, that introducing the same three transcription factors (Brn2, Ascl1/Mash1 and Myt1l) used in their original study into human fibroblasts is not sufficient to convert the cells into functional neurons; the resulting cells are immature, and even after 3 weeks in culture do not become functional, as judged by measurements of action potentials. Nonetheless, the authors show that the efficiency of human

fibroblast conversion into neurons is enhanced by including an extra transcription factor, NeuroD1, a member of the bHLH family.

Caiazzo *et al.*² (page 224) also set out to convert human fibroblasts into a specific type of neuron, in this case neurons of the midbrain that secrete the neurotransmitter dopamine. Starting their study in mice, they tested the efficiency of mixtures of transcription factors (combinations of Ascl1, Brn2, Myt1l and others) in the conversion process. The winning combination was that of Ascl1, Nurr1 and Lmx1a; Brn2 and Myt1l did not lead to dopamine-secreting neurons. The authors provide evidence that the conversion route was direct, and that the resulting neurons were functionally surprisingly similar to primary dopaminergic neurons isolated from the brain. What's more, when the converted dopaminergic cells were grafted into neonatal mouse brains, they became integrated and were functional.

Caiazzo and colleagues then applied their method to human fibroblasts, including those from patients with Parkinson's disease, a neurodegenerative disorder in which dopaminergic neurons are damaged. The conversion efficiency of adult human fibroblasts was 10–20-fold lower than that of their mouse equivalents, and the resulting cells were less mature. This discrepancy points to species differences between mice and humans². The authors' observations fit well with those of

the conversion process is not fully understood. Remodelling of chromatin (DNA–protein complexes), which in turn affects gene expression, seems to have a central role. Yoo *et al.*³ report that the two miRNAs modify the composition of the BAF chromatin-remodelling complex by regulating expression of its subunits, thereby transforming it into a chromatin remodeller that is characteristic of differentiated neurons⁹. But it remains unclear whether the specific chromatin structure in the converted neurons is the same as that in primary neurons.

In this respect, Caiazzo and colleagues' data² are informative. They show that the gene-expression profiles of the dopaminergic neurons generated and those of isolated mouse midbrain dopaminergic neurons are similar but distinct: about 160 genes were expressed differently, with a more than fivefold variation in expression. This indicates that functional similarity of the converted neurons does not necessarily correspond to similar chromatin structure or gene-expression levels, and so cautions against the cells' premature clinical use. Instead, the differences between the converted neurons and their corresponding primary neurons must be characterized, and whether they give rise to unwanted side effects should be explored.

The miRNAs used by Yoo and co-workers³ probably also affect the expression of other proteins involved in cell-fate switching and neuron differentiation, including components of other chromatin-remodelling complexes. So before the clinical use of converted cells can be contemplated, further work should determine how the various chromatin-remodelling factors, and other factors that affect gene expression, contribute to cell conversion and how they can be controlled. Another question to be addressed is how the similarities and differences between converted cells and primary neurons, in terms of gene expression and chromatin structure, correlate with the functionality of the converted neurons — with their neurotransmitter production, firing of action potentials and functional integration into neuronal networks. It is also not known how gene expression and chromatin structure are shaped by intrinsic mechanisms and by the cells' immediate environment, in particular after their transplantation.

Thus, an area for further exploration is how a diseased brain's environment influences the functionality and gene-expression profiles of the transplanted converted neurons. Whether the converted cells are transplanted into the supportive environment of a neonatal mouse brain or into a diseased or aged human brain probably makes a difference. The finding that DNA methylation is dynamic in postnatal neurons¹⁰ raises hopes that the environment into which these converted cells are transferred could contribute to correctly shaping their gene-expression profile for long-term

integration and function. However, the diseased brain in aged humans might not be so good at doing this. Therefore, investigating strategies to compensate for this deficit is perhaps the next big challenge in developing cell-based therapies for neurodegenerative disorders. ■

Michael Sendtner is at the Institute for Clinical Neurobiology, Versbacher Strasse 5, University of Würzburg, 97078 Würzburg, Germany.
e-mail: sendtner_m@klinik.uni-wuerzburg.de

X-RAY IMAGING

The chemistry inside

To understand the properties of many useful materials, the chemical structures that form within them from elements of low relative atomic mass must be determined. A new X-ray imaging technique does just that.

CHRISTIAN G. SCHROER

One of the most striking properties of X-rays is their ability to penetrate matter. In fact, X-rays are partially absorbed as they pass through an object, and cast a 'shadow' of the structures inside — an effect that underpins medical X-ray imaging. As objects become bigger, harder (more energetic) X-rays are needed to penetrate them. The problem is that the harder the X-rays, the smaller the imaging contrast for light elements — those that have low relative atomic masses, such as carbon, nitrogen and oxygen.

Reporting in *Nature Materials*, Huotari *et al.*¹ now describe a way of using hard X-rays to make three-dimensional images of objects that also detects light elements with high sensitivity. The technique even distinguishes between different bonding modes of those elements, and should enable studies of the interiors of diverse objects, such as biological systems, working catalytic reactors and fuel cells.

When a crimson rose is illuminated by sunlight, it appears red to us even though the light contains all the colours of the rainbow. This is because the different colours in the light are absorbed and scattered differently

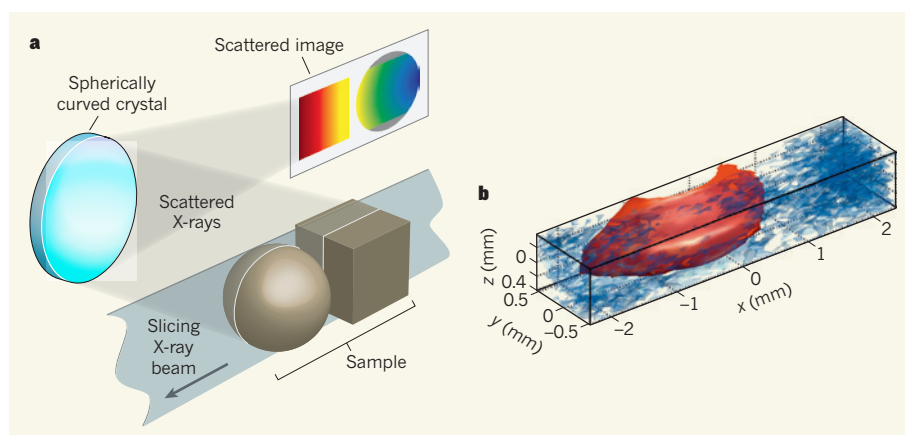


Figure 1 | X-ray scattering for three-dimensional imaging. **a**, Huotari *et al.*¹ report a technique for imaging samples of materials that contain light elements such as carbon, nitrogen and oxygen. A cross-section of a sample is illuminated by a monochromatic, two-dimensional X-ray beam. X-ray scattering occurs, which can alter the energy of the scattered X-rays. A spherically curved crystal filters the scattered X-rays by reflecting only those of a certain pre-set energy, and focuses them onto a detector to form an image of the cross-section. Three-dimensional images are obtained by scanning samples through the incident beam. **b**, The authors' technique can distinguish between different chemical bonding modes in samples. This image shows part of a sample in which a diamond is encased by graphite. Regions containing diamond carbon atoms are shown in red, and those containing graphite carbon atoms are in blue. (Figure adapted from ref. 1.)

the conversion process is not fully understood. Remodelling of chromatin (DNA–protein complexes), which in turn affects gene expression, seems to have a central role. Yoo *et al.*³ report that the two miRNAs modify the composition of the BAF chromatin-remodelling complex by regulating expression of its subunits, thereby transforming it into a chromatin remodeller that is characteristic of differentiated neurons⁹. But it remains unclear whether the specific chromatin structure in the converted neurons is the same as that in primary neurons.

In this respect, Caiazzo and colleagues' data² are informative. They show that the gene-expression profiles of the dopaminergic neurons generated and those of isolated mouse midbrain dopaminergic neurons are similar but distinct: about 160 genes were expressed differently, with a more than fivefold variation in expression. This indicates that functional similarity of the converted neurons does not necessarily correspond to similar chromatin structure or gene-expression levels, and so cautions against the cells' premature clinical use. Instead, the differences between the converted neurons and their corresponding primary neurons must be characterized, and whether they give rise to unwanted side effects should be explored.

The miRNAs used by Yoo and co-workers³ probably also affect the expression of other proteins involved in cell-fate switching and neuron differentiation, including components of other chromatin-remodelling complexes. So before the clinical use of converted cells can be contemplated, further work should determine how the various chromatin-remodelling factors, and other factors that affect gene expression, contribute to cell conversion and how they can be controlled. Another question to be addressed is how the similarities and differences between converted cells and primary neurons, in terms of gene expression and chromatin structure, correlate with the functionality of the converted neurons — with their neurotransmitter production, firing of action potentials and functional integration into neuronal networks. It is also not known how gene expression and chromatin structure are shaped by intrinsic mechanisms and by the cells' immediate environment, in particular after their transplantation.

Thus, an area for further exploration is how a diseased brain's environment influences the functionality and gene-expression profiles of the transplanted converted neurons. Whether the converted cells are transplanted into the supportive environment of a neonatal mouse brain or into a diseased or aged human brain probably makes a difference. The finding that DNA methylation is dynamic in postnatal neurons¹⁰ raises hopes that the environment into which these converted cells are transferred could contribute to correctly shaping their gene-expression profile for long-term

integration and function. However, the diseased brain in aged humans might not be so good at doing this. Therefore, investigating strategies to compensate for this deficit is perhaps the next big challenge in developing cell-based therapies for neurodegenerative disorders. ■

Michael Sendtner is at the Institute for Clinical Neurobiology, Versbacher Strasse 5, University of Würzburg, 97078 Würzburg, Germany.
e-mail: sendtner_m@klinik.uni-wuerzburg.de

X-RAY IMAGING

The chemistry inside

To understand the properties of many useful materials, the chemical structures that form within them from elements of low relative atomic mass must be determined. A new X-ray imaging technique does just that.

CHRISTIAN G. SCHROER

One of the most striking properties of X-rays is their ability to penetrate matter. In fact, X-rays are partially absorbed as they pass through an object, and cast a 'shadow' of the structures inside — an effect that underpins medical X-ray imaging. As objects become bigger, harder (more energetic) X-rays are needed to penetrate them. The problem is that the harder the X-rays, the smaller the imaging contrast for light elements — those that have low relative atomic masses, such as carbon, nitrogen and oxygen.

Reporting in *Nature Materials*, Huotari *et al.*¹ now describe a way of using hard X-rays to make three-dimensional images of objects that also detects light elements with high sensitivity. The technique even distinguishes between different bonding modes of those elements, and should enable studies of the interiors of diverse objects, such as biological systems, working catalytic reactors and fuel cells.

When a crimson rose is illuminated by sunlight, it appears red to us even though the light contains all the colours of the rainbow. This is because the different colours in the light are absorbed and scattered differently

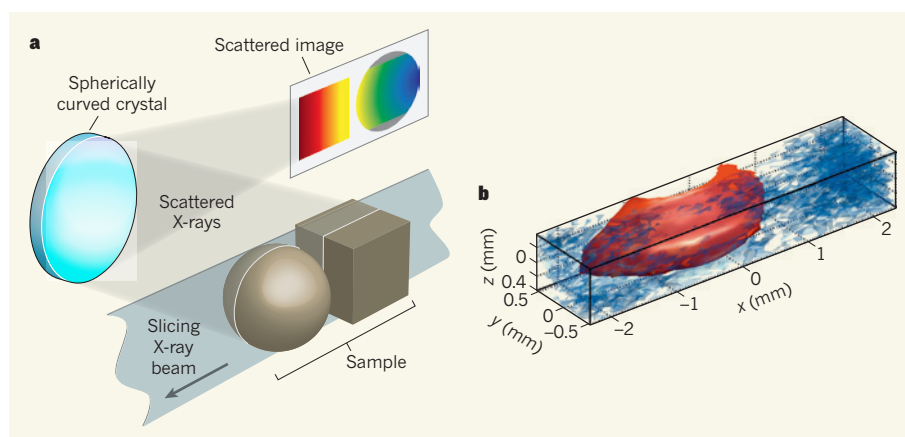


Figure 1 | X-ray scattering for three-dimensional imaging. **a**, Huotari *et al.*¹ report a technique for imaging samples of materials that contain light elements such as carbon, nitrogen and oxygen. A cross-section of a sample is illuminated by a monochromatic, two-dimensional X-ray beam. X-ray scattering occurs, which can alter the energy of the scattered X-rays. A spherically curved crystal filters the scattered X-rays by reflecting only those of a certain pre-set energy, and focuses them onto a detector to form an image of the cross-section. Three-dimensional images are obtained by scanning samples through the incident beam. **b**, The authors' technique can distinguish between different chemical bonding modes in samples. This image shows part of a sample in which a diamond is encased by graphite. Regions containing diamond carbon atoms are shown in red, and those containing graphite carbon atoms are in blue. (Figure adapted from ref. 1.)

by molecules in the petals, which effectively reflect only the red part of the spectrum. Modern spectroscopic techniques make use of this effect to determine the chemical composition and structure of an object. In photoabsorption spectroscopy, visible or ultraviolet light, or X-rays of a given energy, are shone onto an object, and the amount of light absorbed is measured as the radiation passes through². When the energy (colour) of the radiation is varied, the transmittance — a measure of the amount of light that passes through a sample — changes in a way that is characteristic of the atoms in the sample and their chemical bonds to neighbouring atoms. In this way, detailed information about the chemical environments in an object can be obtained.

Carbon, nitrogen and oxygen are particularly abundant in both natural and technologically relevant materials. In order to understand the properties of these materials, it is important to determine the diverse chemical structures that form within them. But to investigate these structures using photoabsorption spectroscopy, for example, soft X-ray light must be used. Most materials are rather opaque to this part of the electromagnetic spectrum — even in water, soft X-rays travel the distance of less than a hair's breadth. It is therefore difficult to use this technique to investigate thicker materials such as biological tissues, or the chemistry in the vicinity of a battery electrode. Analogous electron-based spectroscopic methods could in principle also be used, but electrons penetrate even less deeply into materials than do soft X-rays.

Huotari *et al.*¹ have addressed these difficulties by using hard X-rays, which easily penetrate large sample volumes and a variety of specialized sample environments. In particular, they have used hard X-rays in a technique called inelastic X-ray scattering spectroscopy³. In their system, a sample is partly illuminated by a sheet-like, hard-X-ray beam of fixed energy (Fig. 1a). The X-rays are scattered by the atomic structures in the illuminated slice of the sample. The scattered X-ray light is then reflected by a spherically curved crystal, which focuses a two-dimensional image of the slice onto a detector. Three-dimensional images are obtained by stepping a sample through the beam and assembling the two-dimensional images that are acquired at each step.

In addition to focusing images of samples onto the detector, the curved crystal has a second vital function: because its reflectivity is strongly selective, it reflects and images only scattered light of a certain energy onto the detector. For example, by setting up the system so that only light that has the same energy as the incident beam is reflected, an image can be formed using elastically scattered X-rays — those that are scattered without transferring part of their energy to the sample. Such an image depicts the static (time-averaged) structure of the sample. Similarly, the energy reflected by the crystal can be set to a given

energy transfer (the difference in energy of incident and scattered X-rays). This allows imaging of inelastically scattered X-rays — those that transfer part of their energy to the sample, for example by exciting vibrations, valence electrons or collective motions of electrons.

Crucially for Huotari and colleagues, hard X-rays can also transfer energy to atoms of light elements in excitations similar to those that occur in photoabsorption spectroscopy. The resulting inelastic-scattering spectrum thus reveals the local chemical structure of those atoms in a material. So, by tuning the crystal in their experimental set-up to reflect scattered X-rays at transfer energies corresponding to such excitations, the authors could obtain images that reveal details of the chemical bonds of a given light element in samples.

To illustrate the strength of this approach, Huotari *et al.* analysed a test sample consisting of diamond enclosed by graphite — two allotropes of carbon in which the chemical bonding between atoms is different. Using their technique, the authors were able to map out clearly the positions of the different allotropes in the sample, effectively locating the diamond and graphite components on the basis of their bond structures (Fig. 1b). No other technique could do this, although the diamond could perhaps have been detected using classical absorption imaging because diamond is denser than graphite. But the classical approach would provide no information about the chemical structures of the allotropes.

Because inelastically scattered X-ray

signals are typically weak, the authors' method requires the brightest available X-ray sources. This currently restricts the use of the technique to researchers with access to modern synchrotron radiation sources. Another limitation is the spatial resolution of the technique — the authors studied millimetre-sized samples, but it would be desirable to use the method to study microscopic samples.

Nevertheless, the ability to characterize the chemical bonding of light elements inside thick samples, and of samples in specialized environments (such as in chemical reactors or pressurized cells) opens up many applications. For example, one could use the authors' method to visualize local bond structure in composite materials, or to identify different phases in polymer blends. It could also be used to study the reactions of carbon and oxygen in industrially useful catalytic reactions, such as the catalytic partial oxidation of methane⁴ — one of the steps involved in the conversion of natural gas into liquid fuels. ■

Christian G. Schroer is at the Technische Universität Dresden, Institute of Structural Physics, Dresden 01069, Germany.
e-mail: schroer@physik.tu-dresden.de

1. Huotari, S., Pytkänen, T., Verbeni, R., Monaco, G. & Hämäläinen, K. *Nature Mater.* **10**, 489–493 (2011).
2. Koningsberger, D. C. & Prins, R. (eds) *X-Ray Absorption* (Wiley, 1988).
3. Schülke, W. *Electron Dynamics by Inelastic X-Ray Scattering* (Oxford Univ. Press, 2007).
4. Grunwaldt, J.-D. & Schroer, C. G. *Chem. Soc. Rev.* **39**, 4741–4753 (2010).

GLOBAL CHANGE

The grass response

A three-year study provides insights into how the productivity of a semi-arid rangeland, containing grasses using different photosynthetic pathways, will change in a warmer world with more atmospheric carbon dioxide. [SEE LETTER P.202](#)

DENNIS BALDOCCHI

The grasses that provide forage for most of the world's livestock use either the C₃ or the C₄ photosynthetic routes to fix carbon dioxide into carbohydrates. Both routes eventually employ the carbon-fixing enzyme ribulose biphosphate carboxylase. However, C₄ grasses are more efficient photosynthesizers as a result of their distinct morphology and their use of another enzyme, PEP-carboxylase, before this ultimate step of carbon fixation. On page 202 of this issue, Morgan *et al.*¹ describe experiments aimed at assessing the effects on mixed C₃/C₄ grasslands of raised levels of atmospheric CO₂ and higher temperatures, and, crucially, how these conditions might change the plants' water budgets.

As any high-school science student or climate-change sceptic knows, CO₂ is food for plants. So the presumption that 'more is better', as we release CO₂ at an undiminished rate from fossil-fuel combustion, is partly true. This effect occurs because photosynthesis is stimulated when leaves are exposed to above-ambient levels of CO₂. But the rate of increase in photosynthesis with rising CO₂ will eventually diminish and approach an asymptote as CO₂ levels reach double that of the present day^{2,3}.

How other factors affecting leaf photosynthesis and its conversion of carbohydrates into plant matter respond to elevated CO₂ depends on a variety of conditions: for example, whether plants are growing in isolation or in groups⁴; are woody or herbaceous^{4,5}; are growing as monocultures or as mixed

by molecules in the petals, which effectively reflect only the red part of the spectrum. Modern spectroscopic techniques make use of this effect to determine the chemical composition and structure of an object. In photoabsorption spectroscopy, visible or ultraviolet light, or X-rays of a given energy, are shone onto an object, and the amount of light absorbed is measured as the radiation passes through². When the energy (colour) of the radiation is varied, the transmittance — a measure of the amount of light that passes through a sample — changes in a way that is characteristic of the atoms in the sample and their chemical bonds to neighbouring atoms. In this way, detailed information about the chemical environments in an object can be obtained.

Carbon, nitrogen and oxygen are particularly abundant in both natural and technologically relevant materials. In order to understand the properties of these materials, it is important to determine the diverse chemical structures that form within them. But to investigate these structures using photoabsorption spectroscopy, for example, soft X-ray light must be used. Most materials are rather opaque to this part of the electromagnetic spectrum — even in water, soft X-rays travel the distance of less than a hair's breadth. It is therefore difficult to use this technique to investigate thicker materials such as biological tissues, or the chemistry in the vicinity of a battery electrode. Analogous electron-based spectroscopic methods could in principle also be used, but electrons penetrate even less deeply into materials than do soft X-rays.

Huotari *et al.*¹ have addressed these difficulties by using hard X-rays, which easily penetrate large sample volumes and a variety of specialized sample environments. In particular, they have used hard X-rays in a technique called inelastic X-ray scattering spectroscopy³. In their system, a sample is partly illuminated by a sheet-like, hard-X-ray beam of fixed energy (Fig. 1a). The X-rays are scattered by the atomic structures in the illuminated slice of the sample. The scattered X-ray light is then reflected by a spherically curved crystal, which focuses a two-dimensional image of the slice onto a detector. Three-dimensional images are obtained by stepping a sample through the beam and assembling the two-dimensional images that are acquired at each step.

In addition to focusing images of samples onto the detector, the curved crystal has a second vital function: because its reflectivity is strongly selective, it reflects and images only scattered light of a certain energy onto the detector. For example, by setting up the system so that only light that has the same energy as the incident beam is reflected, an image can be formed using elastically scattered X-rays — those that are scattered without transferring part of their energy to the sample. Such an image depicts the static (time-averaged) structure of the sample. Similarly, the energy reflected by the crystal can be set to a given

energy transfer (the difference in energy of incident and scattered X-rays). This allows imaging of inelastically scattered X-rays — those that transfer part of their energy to the sample, for example by exciting vibrations, valence electrons or collective motions of electrons.

Crucially for Huotari and colleagues, hard X-rays can also transfer energy to atoms of light elements in excitations similar to those that occur in photoabsorption spectroscopy. The resulting inelastic-scattering spectrum thus reveals the local chemical structure of those atoms in a material. So, by tuning the crystal in their experimental set-up to reflect scattered X-rays at transfer energies corresponding to such excitations, the authors could obtain images that reveal details of the chemical bonds of a given light element in samples.

To illustrate the strength of this approach, Huotari *et al.* analysed a test sample consisting of diamond enclosed by graphite — two allotropes of carbon in which the chemical bonding between atoms is different. Using their technique, the authors were able to map out clearly the positions of the different allotropes in the sample, effectively locating the diamond and graphite components on the basis of their bond structures (Fig. 1b). No other technique could do this, although the diamond could perhaps have been detected using classical absorption imaging because diamond is denser than graphite. But the classical approach would provide no information about the chemical structures of the allotropes.

Because inelastically scattered X-ray

signals are typically weak, the authors' method requires the brightest available X-ray sources. This currently restricts the use of the technique to researchers with access to modern synchrotron radiation sources. Another limitation is the spatial resolution of the technique — the authors studied millimetre-sized samples, but it would be desirable to use the method to study microscopic samples.

Nevertheless, the ability to characterize the chemical bonding of light elements inside thick samples, and of samples in specialized environments (such as in chemical reactors or pressurized cells) opens up many applications. For example, one could use the authors' method to visualize local bond structure in composite materials, or to identify different phases in polymer blends. It could also be used to study the reactions of carbon and oxygen in industrially useful catalytic reactions, such as the catalytic partial oxidation of methane⁴ — one of the steps involved in the conversion of natural gas into liquid fuels. ■

Christian G. Schroer is at the Technische Universität Dresden, Institute of Structural Physics, Dresden 01069, Germany.
e-mail: schroer@physik.tu-dresden.de

1. Huotari, S., Pytkänen, T., Verbeni, R., Monaco, G. & Hämäläinen, K. *Nature Mater.* **10**, 489–493 (2011).
2. Koningsberger, D. C. & Prins, R. (eds) *X-Ray Absorption* (Wiley, 1988).
3. Schülke, W. *Electron Dynamics by Inelastic X-Ray Scattering* (Oxford Univ. Press, 2007).
4. Grunwaldt, J.-D. & Schroer, C. G. *Chem. Soc. Rev.* **39**, 4741–4753 (2010).

GLOBAL CHANGE

The grass response

A three-year study provides insights into how the productivity of a semi-arid rangeland, containing grasses using different photosynthetic pathways, will change in a warmer world with more atmospheric carbon dioxide. [SEE LETTER P.202](#)

DENNIS BALDOCCHI

The grasses that provide forage for most of the world's livestock use either the C₃ or the C₄ photosynthetic routes to fix carbon dioxide into carbohydrates. Both routes eventually employ the carbon-fixing enzyme ribulose biphosphate carboxylase. However, C₄ grasses are more efficient photosynthesizers as a result of their distinct morphology and their use of another enzyme, PEP-carboxylase, before this ultimate step of carbon fixation. On page 202 of this issue, Morgan *et al.*¹ describe experiments aimed at assessing the effects on mixed C₃/C₄ grasslands of raised levels of atmospheric CO₂ and higher temperatures, and, crucially, how these conditions might change the plants' water budgets.

As any high-school science student or climate-change sceptic knows, CO₂ is food for plants. So the presumption that 'more is better', as we release CO₂ at an undiminished rate from fossil-fuel combustion, is partly true. This effect occurs because photosynthesis is stimulated when leaves are exposed to above-ambient levels of CO₂. But the rate of increase in photosynthesis with rising CO₂ will eventually diminish and approach an asymptote as CO₂ levels reach double that of the present day^{2,3}.

How other factors affecting leaf photosynthesis and its conversion of carbohydrates into plant matter respond to elevated CO₂ depends on a variety of conditions: for example, whether plants are growing in isolation or in groups⁴; are woody or herbaceous^{4,5}; are growing as monocultures or as mixed

species^{4,5}; are growing under natural or managed conditions^{4–6}; utilize the C₃ or the C₄ photosynthetic pathway^{3,7}; are evaluated after the course of a growing season or after many years^{4,5,8}; or are exposed to a warmer or drier climate³.

Morgan *et al.*¹ disentangle part of this web of potentially interacting factors by presenting convincing evidence that, compared with C₃ grasses, C₄ grasses growing under semi-arid climatic conditions will prosper in a world with higher CO₂ concentrations and warmer temperatures. In principle, this response depends on water — the leaf pores, known as stomata, of C₄ plants open less widely than those of C₃ plants, so C₄ plants tend to lose less water vapour through transpiration.

Morgan and colleagues' take-home message is the following: water savings proffered to C₄ grasses from partial stomatal closure, induced by high CO₂ concentrations, offset any enhancement in evaporation brought about by coincident warming. They draw this conclusion from a novel, multifactorial experiment that exposed plots of C₃ and C₄ grasses to combinations of ambient or elevated CO₂ and/or temperature for three years. For the single-factor plots, they found that elevated CO₂ favours the productivity of C₃ grasses and that elevated temperature favours that of C₄ grasses, as expected³. However, a combination of elevated CO₂ and higher temperature favours increased C₄ productivity because water-use efficiency (the ratio between photosynthesis and transpiration) in C₄ grasses is greater than in C₃ grasses.

Thousands of papers and hundreds of reviews⁴ have been published on how plants respond to elevated CO₂. So why is this particular study noteworthy?

First, Morgan and colleagues¹ provide one of the first and best views of how a mixed-grass ecosystem (Fig. 1) growing in a semi-arid climate will respond to future CO₂ and climatic conditions. The task of divining this information is not trivial, because the perturbation of an ecosystem engenders a suite of nonlinear, positive and negative feedbacks that operate across an array of timescales⁴. Second, their data counter the conventional wisdom derived from physiological studies that leaf photosynthesis and water-use efficiency in C₄ grasses benefit less from high CO₂ than in C₃ grasses³. Third, the interactive effects of CO₂, temperature and soil moisture on C₃ and C₄ grasses have until now remained elusive. Without the warming treatment, Morgan and colleagues' results would only have confirmed knowledge gained from previous meta-analyses that compared the responses of C₄ and C₃ grasses to raised CO₂ — both C₃ and C₄ plants respond positively to enhanced CO₂ (ref. 7).

However, users of this new information¹ must appreciate the conditional aspects of Morgan and colleagues' results. Short-term studies of elevated CO₂, over single



Figure 1 | Mixed-grass prairie. This grassland in Wyoming is covered by a variety of plants that use either C₃ or C₄ photosynthetic routes, and is the kind of ecosystem studied by Morgan and colleagues¹.

to several growing seasons, manipulate the ecosystem's eco-physiological 'knobs', such as stomatal conductance, leaf-area index and the soil–water balance. In such circumstances, elevated CO₂ will indeed cause stomata to close partially, restricting leaf transpiration and conserving soil moisture. But enhanced growth increases the population of leaves and so counteracts potential soil-moisture savings — up to a limit.

Whether this net moisture saving is critical to plant productivity depends on how much rain falls during the growing season. In a more humid environment, the water saving through CO₂-induced stomatal closure in C₄ plants may be less consequential, as shown in 2009 when precipitation at Morgan and colleagues' test site was greater than normal¹. If there is severe drought, however, a rise in CO₂ may not provide enough water savings to make a difference. When warming is added to the mix, it can counter the leaf's physiological response to elevated CO₂. Concurrent warming promotes transpiration by increasing the gradient in humidity between the leaf surface and the atmosphere. Warming also increases rates of leaf and root respiration, thereby diminishing net carbon acquisition and reducing plant growth. Consequently, the answer to the question, 'How will this mixed-grass ecosystem respond to changes in mean CO₂?' will depend, in part, on whether or not climatic extremes in the moisture and thermal environments occur and physiological thresholds are crossed.

In contrast to short-term experiments, studies lasting longer than a decade manipulate biogeochemical and ecological knobs such as leaf and root nutrition, species composition and plant competition. Downregulation of photosynthesis typically occurs in long-term studies of elevated CO₂, because the nitrogen needed for ribulose biphosphate carboxylase becomes scarcer as plant biomass expands and

sequesters the limited supply of nitrogen available to the plants' roots^{2,5,8}. These longer-term responses are not captured in Morgan and colleagues' three-year study. Hence, additional years of studying this ecosystem may produce alternative outcomes. Moreover, findings that apply to mixed C₃/C₄ grasslands may not relate to other ecosystems such as forests⁴.

What do these results¹ suggest for the future? At present, C₄ grasses tend to inhabit warmer and drier climates than do C₃ grasses^{3,7}. If the trends in atmospheric warming and increasing CO₂ persist over decades, a gradual shift in species composition may occur, favouring C₄ grasses. Any change in the mixture of C₃ and C₄ grasses will have a biophysical feedback on the climate system by changing how solar energy is absorbed and partitioned into heating the air and evaporating soil moisture⁹. Such changes in species composition may also affect ecosystem services — for example, the amount and availability of forage for grazing cattle, bison and antelope. ■

Dennis Baldocchi is in the Department of Environmental Science, Policy and Management, and the Berkeley Atmospheric Sciences Center, University of California, Berkeley, Berkeley, California 94720-3110, USA. e-mail: baldocchi@berkeley.edu

1. Morgan, J. A. *et al.* *Nature* **476**, 202–205 (2011).
2. Ainsworth, E. A. & Rogers, A. *Plant Cell Environ.* **30**, 258–270 (2007).
3. Pearcy, R. W. & Ehleringer, J. *Plant Cell Environ.* **7**, 1–13 (1984).
4. Körner, C. *New Phytol.* **172**, 393–411 (2006).
5. Ainsworth, E. A. & Long, S. P. *New Phytol.* **165**, 351–372 (2005).
6. Leakey, A. D. *et al.* *J. Exp. Bot.* **60**, 2859–2876 (2009).
7. Wand, S. J. E., Midgley, G. F., Jones, M. H. & Curtis, P. S. *Glob. Change Biol.* **5**, 723–741 (1999).
8. Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E. & McMurtrie, R. E. *Proc. Natl Acad. Sci. USA* **107**, 19368–19373 (2010).
9. Bounoua, L., DeFries, R., Collatz, G. J., Sellers, P. & Khan, H. *Clim. Change* **52**, 29–64 (2002).

A continuum model for tumour suppression

Alice H. Berger¹, Alfred G. Knudson² & Pier Paolo Pandolfi¹

This year, 2011, marks the forty-year anniversary of the statistical analysis of retinoblastoma that provided the first evidence that tumorigenesis can be initiated by as few as two mutations. This work provided the foundation for the two-hit hypothesis that explained the role of recessive tumour suppressor genes (TSGs) in dominantly inherited cancer susceptibility syndromes. However, four decades later, it is now known that even partial inactivation of tumour suppressors can critically contribute to tumorigenesis. Here we analyse this evidence and propose a continuum model of TSG function to explain the full range of TSG mutations found in cancer.

Although hereditary predisposition to cancer was known before 1900, it was only then, after the rediscovery of Gregor Mendel's once-ignored nineteenth century work, that hereditary predisposition to cancer could be rationalized. By then it was also known that the pattern of chromosomes in cancer cells is abnormal. The next contribution to understanding cancer genetics was made by Theodor Boveri, who proposed that some chromosomes might stimulate cell division and others might inhibit it, but his idea was long overlooked. Now we know that there are genes of both types.

In this review, we summarize the history of the study of the latter type of genes, tumour suppressor genes (TSGs), and the evidence supporting a role for both complete and partial tumour suppressor inactivation in the pathogenesis of cancer. We integrate the classical 'two-hit' hypothesis of tumour suppression with a continuum model that accounts for subtle dosage effects of tumour suppressors, and we discuss other exceptions to the two-hit hypothesis such as the emerging concept of 'obligate haploinsufficiency' in which partial loss of a TSG is more tumorigenic than complete loss. The continuum model highlights the importance of subtle regulation of TSG expression or activity such as regulation by microRNAs (miRNAs). Finally, we discuss the implications of this model for the diagnosis and therapy of cancer.

The two-hit hypothesis

The first evidence for a genetic abnormality as a cause of cancer came with the discovery, in 1960, of the abnormal 'Philadelphia' chromosome in chronic myelogenous leukaemia cells¹. Later, in 1973, it was discovered that this chromosome was a translocation between chromosomes 9 and 22 (ref. 2) and in 1977, a translocation of chromosomes 15 and 17 was identified in acute promyelocytic leukaemia³. Eventually, the genes at the break-points of these translocations were cloned: *BCR/ABL1* in 1983 (ref. 4) and *PML/RARA* in 1991 (ref. 5). Meanwhile, seminal work demonstrated that the cancer-causing gene within the avian sarcoma virus genome, *v-src*, was actually a co-opted and mutated version of a normal cellular gene termed a proto-oncogene, now called *c-src*^{6,7}. These observations demonstrated that normal cellular genes, when mutated or altered, are able to cause cancer, and were followed by the identification of numerous other cellular oncogenes that are activated by mutation, chromosome translocation or amplification. One of Boveri's predictions was proven to be correct.

In the 1980s, scientists verified Boveri's other prediction with the identification of a second class of genes involved in cancer, tumour suppressor

genes, that inhibit cancer development and oppose oncogene function. In a normal cell, a physiological balance between tumour suppressors and oncogenes maintains homeostasis and allows carefully regulated cell proliferation without unrestrained malignant tumour growth. Somatic cell fusion experiments pointed to the existence of such TSGs, because fusion of a normal cell with a malignant cell could revert the malignant cell to a normal phenotype⁸. One paradox that puzzled cancer researchers was that cancer susceptibility syndromes usually display a dominant mode of inheritance, whereas the proposed tumour suppressor genes seemed to function in a recessive manner in the *in vitro* cell fusion experiments.

The resolution of this problem was provided by the analysis of a tumour of children, retinoblastoma, which is sometimes present even at birth⁹. Statistical modelling indicated that hereditary cases probably developed after only one somatic mutational event. It was proposed that one mutant allele is inherited and the other is generated somatically during growth of the developing eye. Because of the high chance of the second 'hit' occurring, almost all individuals with the inherited first hit (the mutant allele) will develop retinoblastoma, and therefore the cancer-susceptibility phenotype is inherited in a dominant manner (Fig. 1). In contrast, tumour initiation requires both hits, and therefore tumorigenesis is recessive. A similar mechanism could be extended to sporadic retinoblastoma, the only difference being that in the sporadic cases both the first and second hits occur in somatic tissue. At the time of the original analysis, it was not known whether these two hits occurred in the same or distinct genes, but it was clear that the search for the secondary mutation would start with the disease gene itself.

The success of the two-hit model came with the identification of deletion of a locus on chromosome 13 in retinoblastoma tumours and the subsequent cloning of the *RBI* gene as mutated in familial retinoblastoma^{10–17}. This discovery was made possible by the development of restriction fragment length polymorphism (RFLP) mapping technology. Using this approach, it was shown that many families with familial retinoblastoma did indeed harbour mutations or deletions of *RBI* in the germ line (the first hit), and tumours from retinoblastoma patients nearly always contained mutation or loss of the other *RBI* allele^{10–17} (the second hit). This latter event most often results from somatic mitotic recombination, whereas in a minority of cases, the normal allele acquires a distinct mutation. The role of recessive TSGs in dominantly inherited cancer susceptibility syndromes could then be understood, and Boveri's second prediction was substantiated.

¹Cancer Genetics Program, Beth Israel Deaconess Cancer Center, Departments of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, USA.

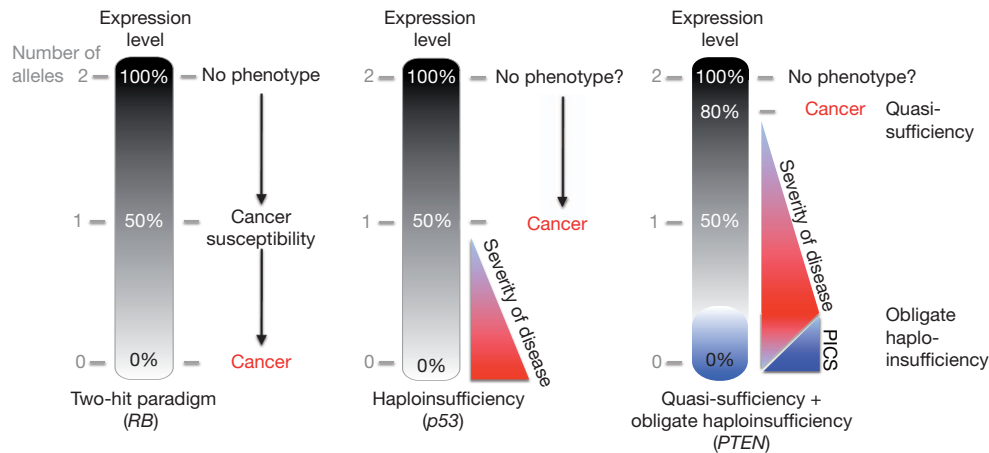


Figure 1 | Paradigms of tumour suppression. A black gradient represents a continuum of expression related to the number of alleles present (grey numbering). Left, the two-hit paradigm as exemplified by the tumour suppressor, *RB*. Loss of one allele induces cancer susceptibility; loss of two alleles induces cancer. Middle, classical haploinsufficiency. Loss of one allele is sufficient for induction of cancer. Right, quasi-sufficiency and obligate

haploinsufficiency. Quasi-sufficiency refers to the phenomenon whereby tumour suppression is impaired after subtle expression downregulation without loss of even one allele. Obligate haploinsufficiency occurs when TSG haploinsufficiency is more tumorigenic than complete loss of the TSG, usually due to the activation of fail-safe mechanisms following complete loss of TSG expression.

These findings and others not only established the two-hit hypothesis as a model that explained hereditary cancer susceptibility, but also provided a proof-of-principle that TSGs could be identified by the study of chromosomal deletions and genetic linkage in hereditary cases of a cancer. Such a strategy was critical to the ultimate identification of *TP53* (also called *p53*) as a TSG¹⁸ and was used successfully in the identification of other TSGs such as *APC*¹⁹, *BRCA1* (ref. 20) and *BRCA2* (ref. 21). Now the advent of high-throughput sequencing and genome-wide copy number profiling is allowing further systematic identification of regions of mutation and deletion in tumour genomes^{22,23}.

It should be noted that it was initially proposed that two hits would be sufficient for the initiation of some tumours such as paediatric cancers, but additional mutations would be necessary for others. It is now thought that non-hereditary cancers require approximately four or more distinct mutation events that result in perturbation of critical cellular signalling pathways such as phosphoinositide-3-kinase (PI(3)K), *p53* or *RB*²⁴. Malignancy is thought to result from an iterative process of somatic mutation followed by clonal expansion²⁵. In the context of hereditary cancer syndromes, full inactivation of the involved disease gene may be the rate-limiting step for tumour initiation, but other events may occur, particularly to promote or enhance tumour progression. In this respect, the two-hit hypothesis should be applied directly to the TSG, with two hits representing the number of events necessary for inactivation of the TSG, but not necessarily for tumorigenesis. Haploinsufficiency, as we discuss below, must also be interpreted in this context; in many cases, a gene might be haploinsufficient for a specific cellular function, but TSG haploinsufficiency may promote tumour formation only in the context of other genetic or environmental insults.

Haploinsufficiency and quasi-sufficiency

The two-hit hypothesis clearly explained the differences in tumour number and age of onset of retinoblastoma in hereditary versus sporadic cases of retinoblastoma. Moreover, it led to the eventual cloning of *RBI* as the first bona fide human TSG. The hypothesis has been useful in cloning genes for hereditary cancers and for their non-hereditary counterparts. However, this success has led to a questionable dogma in the field that all important tumour suppressors should behave according to the two-hit model.

Investigators sought to apply the principles learned from the two-hit model to sporadic cancers in order to identify novel TSGs. Many recurring regions of chromosomal deletion have been identified in sporadic cancers, indicative of the presence of a TSG. Unfortunately, it has become apparent that not all regions of consistent loss are accompanied by obvious aberrations on the other allele, raising the question of

whether these deletions may represent 'passenger' events not relevant to carcinogenesis. Indeed, very few of the identified regions of chromosomal losses have yielded clear TSGs in those loci that conform to the two-hit hypothesis. This finding has led many to conclude that the relevant tumour suppressors in those regions have not been identified or do not exist, when in fact many genes in those loci are known to have tumour suppressive properties *in vitro* or *in vivo*.

Alternatively, these regions could harbour bona fide TSGs but may represent regions/genes where single-copy mutation or loss has a role in tumorigenesis. These single-copy events may even be selected for during tumorigenesis instead of biallelic TSG loss. One possibility is that the initial lesion, when reduced to homozygosity, may lead to cell death or senescence. The lethality could be due to homozygosity of the disease gene itself or of other distinct genes included in the initial hit, because recombination is a principal second hit in tumorigenesis²⁶. Such a scenario would result in an obligate haploinsufficiency in which selection pressure during tumorigenesis favours partial, but not complete, loss of the TSG.

A second possibility is that the single-copy mutation of a TSG functions as a dominant-negative mutation towards the wild-type gene/protein. After mutation of one allele, the mutant protein product interferes with the normal wild-type protein produced from the remaining wild-type allele. Because the complete normal function of the TSG is already impaired with only one hit, there is no selection pressure by the tumour for loss or mutation of the wild-type allele.

A third possibility is that a gene or genes in the region of consistent deletion shows haploinsufficiency for tumour suppressor function and single-copy loss of the TSG is sufficient for aberrant TSG function and promotion of cancer (Fig. 1). In contrast with classical TSGs that are insensitive to large reductions in expression or activity, the function of haploinsufficient TSGs is impaired by a 50% partial reduction in expression or activity, or sometimes by subtler changes, as in a quasi-insufficiency of TSG function that we discuss below.

The role of haploinsufficiency in cancer has been met with scepticism, despite the well-established role of haploinsufficiency in numerous developmental disorders such as, for instance, aniridia (caused by haploinsufficiency for *PAX6*) and Grieg's syndrome (caused by haploinsufficiency of *GLI3*)²⁷. Just as increased dosage of genes can result in developmental syndromes, such as Down's syndrome, increased dosage or activity of oncogenes is an established genetic mechanism of malignancy²⁸, exemplified by the role of *MYC* amplification or *RAS* protein hyperactivity in cancer. However, the notion that subtle decreases in gene dosage or protein activity can be relevant to cancer has gained only limited acceptance in the scientific community.

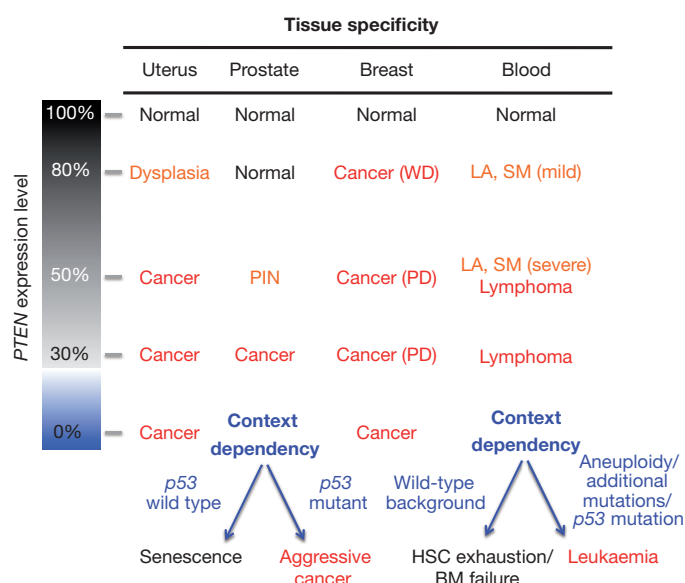


Figure 2 | Tissue specificity and context dependency of tumour suppression. The phenotypic outcome of a reduction in *PTEN* expression is differentially manifested depending on tissue type and genetic background. BM, bone marrow; HSC, haematopoietic stem cell; LA, lymphadenopathy; PD, poorly differentiated; SM, splenomegaly; WD, well-differentiated. The effect of complete loss of *PTEN* is highly context-dependent due to the obligate haploinsufficiency caused by *PTEN* loss-induced cellular senescence (PICS). Data summarized here come from multiple groups and studies of genetically engineered mice with differing *PTEN* alleles and expression^{38–40,52,65–68,70,71}.

One reason for this scepticism stems from the difficulty in definitively proving that a haploinsufficient TSG is involved in tumorigenesis. Whereas the rare TSGs that fully conform to the two-hit hypothesis can be identified by their homozygous deletion or mutation in cancer, haploinsufficient TSGs cannot be identified in such a manner. Moreover, large regions of the genome, encompassing many genes, can be targeted by allelic loss and likewise, many genes may acquire somatic mutations during the course of tumour development. It is assumed that only a portion of these genes is responsible for the cancer, while the rest are passenger mutations. What then could be the gold standard by which to determine which gene or genes in the region are causative and which are non-involved bystanders? Unlike the analysis of traditional TSGs, no one assay or approach offers definitive proof that a haploinsufficient TSG is causally involved in tumorigenesis. Instead, an integrated approach involving human genetic analysis, *in vitro* and *ex vivo* functional studies, and murine cancer genetic modelling is critical to ascertain whether a gene may function as a haploinsufficient gene in cancer. Although there are caveats of each analysis on its own, a body of evidence from each of these analyses can strongly implicate a gene in dosage-sensitive tumour suppression.

Throughout this review, we use *PTEN* as a model dosage-sensitive TSG, but numerous other TSGs exhibit haploinsufficiency and dosage-sensitivity. Notably, *p53* shows haploinsufficiency. Mice with heterozygous *p53* mutation (*p53*^{+/-}) show an intermediate survival to that of *p53* homozygous mutants and wild-type mice, and tumours that develop in the *p53*^{+/-} animals do not always display loss of the remaining wild-type allele²⁹. Similarly, tumours in patients with Li–Fraumeni syndrome, a cancer susceptibility syndrome caused by germ line mutation of *p53*, do not always exhibit loss of the wild-type *p53* allele, suggesting that haploinsufficiency of *p53* may be sufficient for tumour initiation in humans as well³⁰. A caveat with these analyses is that the effect of *p53* haploinsufficiency on cancer initiation in mouse models and Li–Fraumeni patients may be due to complex non-cell autonomous effects stemming from the partial loss of *p53* throughout the body. However, *ex vivo* analysis of murine *p53* heterozygous thymocytes showed that these

cells have an impaired apoptotic response to ionizing radiation or etoposide³¹. Similarly, a *p53*-deleted HCT116 isogenic cell line expressed only 25% the normal level of *p53* messenger RNA and showed impairment in induction of *p53*-responsive genes and apoptosis after exposure to ultraviolet radiation³². These studies demonstrate that reductions in *p53* dosage and function can have an impact on a cell's ability to respond to oncogenic stimuli and strengthen the notion that *p53* haploinsufficiency may drive tumorigenesis.

Similarly, other classical cancer susceptibility genes such as *BRCA1* and *BRCA2* may show haploinsufficiency and/or dosage effects. Primary breast and ovarian cells from patients with heterozygous germ line mutation of *BRCA1* or *BRCA2* have altered mRNA profiles compared to *BRCA* wild-type cells, indicating that single hits in these genes can confer phenotypic differences that could, in principle, have an effect on breast and/or ovarian tumorigenesis³³. Indeed, *ex vivo* studies of human *BRCA1* mutant breast cells have found that these cells show enhanced colony formation potential³⁴ and show impaired lineage commitment in differentiation assays³⁵. Although tumour formation in *BRCA1* and *BRCA2* carriers does appear to require the second hit, these one-hit premalignant changes could have an impact on tumorigenesis by either promoting loss of the second TSG allele or of other TSGs, or by altering the cellular phenotype to predispose to certain tumour subtypes.

Another haploinsufficient TSG is the transcription factor *PAX5*. *PAX5* is found mutated, deleted or translocated in approximately 30% of acute lymphoblastic leukaemia cases, but the aberrant mutations/deletions are almost invariably mono-allelic and seem to function as hypomorphs, not dominant negative alleles³⁶. Similarly, the E3 ligase gene *FBW7* (also known as *FBXW7*) is mutated in approximately 10% of human colorectal cancers³⁷. In 70% of these cases, the mutations are mono-allelic, indicative of haploinsufficiency. Conditional mono-allelic deletion of *Fbw7* in the murine intestine cooperated with the tumorigenic *APC*^{min} allele in intestinal tumorigenesis, albeit to a lesser degree than biallelic deletion of *Fbw7*, indicating that *Fbw7* is dosage-sensitive³⁷.

Like complete loss of TSGs, the effect of TSG haploinsufficiency can be highly tissue-specific and context-dependent (Fig. 2). Thus, the cellular and molecular context in which TSG function is altered will determine the outcome of such impairment of TSG function. The differing outcomes of TSG expression level in differing situations may reflect distinct thresholds of protein expression or activity needed for certain processes or in differing cell contexts. For example, in one cell type there may be other compensatory proteins that mask the potential phenotype caused by TSG haploinsufficiency, whereas in other cells these proteins are not expressed and so the haploinsufficiency of the TSG manifests as tissue-specific cancer susceptibility.

Two hits for tumour suppression

The genetic background of the individual or tumour will also have an impact on the phenotypic outcome of TSG haploinsufficiency. In some cases, combinatorial effects could arise whereby haploinsufficiency or mutation of other genetic loci is required for a phenotype caused by TSG haploinsufficiency. A powerful example of this context dependency is provided by the interaction between *Pten* deficiency or haploinsufficiency and the presence or absence of mutated *p53* in mouse models (ref. 38; Fig. 2). In the context of wild-type *p53*, *Pten* haploinsufficiency is actually more tumorigenic in the prostate than complete loss of *Pten*, because the latter triggers a *p53*-dependent fail-safe senescence mechanism called *Pten*-loss-induced cellular senescence (PICS). Haploinsufficiency of *Pten* enhances proliferation without activating PICS, making partial loss of *Pten* more detrimental in this context than complete loss of *Pten* (Fig. 2). Similarly, complete loss of *Pten* in the murine haematopoietic compartment induces haematopoietic stem cell exhaustion or bone marrow failure, unless loss of *Pten* is accompanied by other genetic events³⁹. After accumulation of aneuploidy or other mutations, such as loss of *p53*, complete loss of *Pten* in the blood induces fatal leukaemias^{39,40} (Fig. 2).

This phenomenon defines a novel paradigm of obligate haploinsufficiency in which selection pressure during tumour growth may favour

haploinsufficiency of a TSG over complete loss in certain scenarios (Fig. 1). However, in advanced cancers with *p53* mutation or loss, PICS cannot be induced and complete loss of *Pten* enhances proliferation and tumorigenesis to a greater degree than *Pten* haploinsufficiency. This remarkable finding is likely to explain why complete loss of *PTEN* is typically restricted to advanced cancers and also emphasizes the importance of single-hit haploinsufficiency of TSGs in cancer initiation. Understanding the combinatorial and contextual dependencies of genetic events should not only enhance our ability to effectively treat tumours with differing combinations of genetic hits, but will also inform chemopreventative strategies for eradicating cancer by attacking mechanisms necessary for its development.

Like *PTEN*, many other TSGs have obligate haploinsufficiency. One example is the microRNA-processing enzyme *DICER1*. Mono-allelic deletion or mRNA downregulation of *DICER1* is observed in diverse cancers^{41,42}. In murine models, mono-allelic conditional deletion of *Dicer1* enhances lung tumorigenesis driven by oncogenic *Kras*⁴¹, sarcoma development induced by *Kras*^{G12D} and loss of *p53* (ref. 41), and retinoblastoma formation induced by deletion of *Rb1* and *p107* (ref. 42). In the lung model, complete inactivation of *Dicer1* improved survival compared to animals with mono-allelic *Dicer1* deletion⁴¹. Tumours that did develop in the *Dicer1*^{fllox/fllox} animals retained one wild-type allele and partial *Dicer1* expression, strongly suggesting that partial loss of *Dicer1* promotes tumorigenesis.

A recent *in vivo* short hairpin RNA (shRNA)-based screening approach in mice identified lymphoma tumour suppressors, including one, the cell cycle regulator *Rad17*, which seemed to function as a TSG with obligate haploinsufficiency⁴³. The use of multiple hairpins targeting *Rad17* led to the discovery that only hairpins that partially suppressed *Rad17* promoted lymphomagenesis, whereas hairpins that completely repressed *Rad17* expression were selected against in the lymphomas. This study highlights the utility of shRNA knockdown and screening technologies for the identification of haploinsufficient TSGs.

NPM1, found mutated and translocated in haematopoietic malignancies, also shows obligate haploinsufficiency. Whereas single-allelic deletion of *Npm1* induces tumour formation in mice⁴⁴, biallelic loss induces embryonic lethality and is also incompatible with cell growth after conditional deletion *ex vivo*⁴⁵. Although more data are needed the tumour-suppressive isoform of *Trp63*, *Tap63*, may also show obligate haploinsufficiency in mice; heterozygosity for *Tap63* promoted tumorigenesis to a greater degree than complete loss of *Tap63* in genetically engineered mice, and tumours that developed in the heterozygous mice did not display loss of the wild-type allele⁴⁶. Similarly, *PinX1*, recently identified as a haploinsufficient tumour suppressor⁴⁷, may also show obligate haploinsufficiency; heterozygous loss of *PinX1* in a mouse model promoted cancer development in multiple organs, and tumours did not lose the wild-type allele or expression of *PinX1* (ref. 47).

A genome-wide screening study in yeast indicated that about 3% of the genes in the yeast genome showed haploinsufficiency when assayed under specific growth conditions⁴⁸. These haploinsufficient genes were enriched for ribosomal genes and genes that are highly expressed. Currently, mammalian tumour suppressors that show haploinsufficiency do not seem to conform to one specific functional class. Besides *p53* and *PTEN*, dozens of other TSGs show haploinsufficiency for tumour suppression^{49,50}, many of which we have discussed above. These TSGs come from many functional classes, including transcription factors (for example, *PAX5* and *Tap63*), cell cycle checkpoint genes (*RAD17*), E3 ligases (*FBW7*), ribosomal or associated proteins (*NPM1*), and other diverse genes such as *DICER1* and *PINX1*.

A continuum model for tumour suppression

Rather than TSG function and activity following discrete, step-wise changes, evidence is emerging that supports the dosage-dependency of TSG function. Data from an allelic series of targeted *Pten* alleles that results in varying levels of *Pten* *in vivo* in genetically engineered mice^{51,52} demonstrate the tight correlation between *Pten* expression level and

function. Taking advantage of a null *Pten* allele in combination with a *Pten* 'hy' allele, which is expressed at a lower level than the wild-type allele, an allelic series was created in which *Pten* expression follows a decreasing pattern in the order *Pten*^{+/+} > *Pten*^{hy/+} > *Pten*^{+/-} > *Pten*^{hy/-}. Surprisingly, *Pten*^{hy/+} animals, with cells expressing 80% of the normal level of *Pten*, did display many of the phenotypes observed in *Pten*^{+/-} animals, such as lymphadenopathy, splenomegaly and mammary gland tumours⁵². Moreover, survival of *Pten*^{hy/+} animals is compromised compared to *Pten*^{+/+} controls, but not to the same degree as the survival impairment in *Pten*^{+/-} animals⁵². Thus, the tumour incidence and survival of the animals correlate with the *Pten* expression level⁵². Critically, mammary tumours from *Pten*^{hy/+} animals do not show loss of *Pten* expression or loss of the wild-type *Pten* allele and exhibit a *Pten* dose-dependent increase in activation of Akt, further demonstrating that a subtle reduction in *Pten* dose can result in deregulated signalling and cancer⁵². Similarly, in the murine prostate, a progressive lowering of *Pten* levels correlates with Akt activation, prostate enlargement, proliferation of prostate cells and tumorigenesis⁵¹. These *in vivo* studies demonstrate that even a subtle 20% reduction of protein level could constitute an important hit involved in the development of cancer. We termed this state: TSG 'quasi-insufficiency'⁵² (Fig. 1).

Thus, protein function can be a continuum that is related to the level of expression or activity of the TSG rather than to discrete step-by-step changes in gene copy number. We therefore propose a shift from the classical discrete model of tumour suppression to a continuum model (Fig. 3). This model takes into account the fact that subtle regulation of TSGs can have profound consequences on cancer susceptibility and/or progression, and in turn has important implications for cancer diagnosis and therapy. Note that whereas the function of *PTEN* seems to be strongly coupled to its expression level, for other TSGs it will be the activity of the protein and not the expression level that may be critical. This situation is readily observed in tumours with *p53* mutation where the mutant allele is highly expressed but functionally inactive or acting as a dominant negative.

Like the effect on TSGs, dosage can be extremely important for regulating the cellular effects of oncogene expression. Markedly different

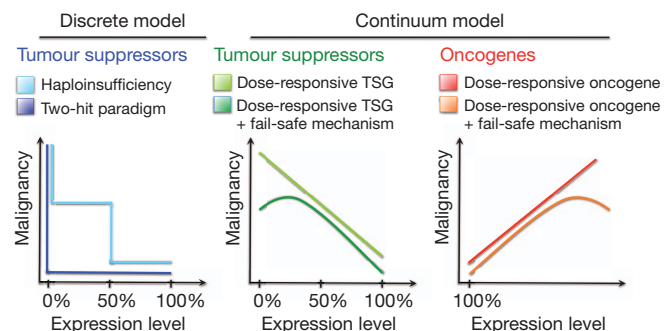


Figure 3 | The continuum model of tumour suppression. The classical discrete, step-wise model of tumour suppression (left) is contrasted with a continuum model of tumour suppression and oncogenesis (centre and right, respectively). In the discrete model, tumorigenesis is induced by either complete loss of a TSG (two-hit paradigm, dark blue) or after single-copy loss of a TSG (haploinsufficiency, light blue). In contrast, we propose a continuum model (centre and right), in which tumour suppression is related to a continuum of TSG expression, rather than to discrete changes in DNA copy number. A continuum of increasing TSG expression will generally be negatively correlated with malignancy (centre, light green), whereas increasing oncogene expression will generally be positively correlated with malignancy (right, red). A linear relationship is depicted for schematic purposes, but the dose-response relationship need not be linear. In some cases, fail-safe mechanisms are induced by complete loss of TSG expression or by massive oncogene overexpression. In these cases, complete loss of TSG expression (centre, dark green) or massive overexpression of an oncogene (right, orange) will be negatively correlated with malignancy, as shown.

consequences for the cell and for cancer development can occur depending on the level of oncogene expression. For example, high expression of oncogenic, mutation of *RAS*-family genes drives senescence, not proliferation⁵³. Similarly, robust expression of *MYC* induces apoptosis, not proliferation⁵⁴. These examples mirror the paradigm of PICS and obligate haploinsufficiency and suggest that tumours may select for optimal expression of oncogenes. Data to support this notion exist already in the rampant amplification of mutated oncogenes, such as the amplification of the mutated *EGFR* gene in human lung cancer^{55,56}. Such data indicate that it is not only the mutation of an oncogene that is important for cancer, but also its expression level. In this respect, the continuum model could also apply to oncogenes (Fig. 3).

It seems now that some TSGs are exquisitely sensitive to dose (for example, *PTEN*), some are intermediately sensitive (for example, *p53*) and others are more resistant to expression level and/or activity changes (for example, *RB1*). Those that are most resistant to changes in expression or activity, such as *RB1*, will be most likely to conform to the two-hit hypothesis, whereas those that are exquisitely sensitive to dose may exhibit frequent monoallelic deletion/mutation or be misregulated in cancer via other means, such as miRNA overexpression. Moreover, tumour suppressors that show obligate haploinsufficiency due to the lethality of homozygous inactivation of the TSG will never show complete inactivation in cancer unless there are other context- or genotype-dependent events that favour their complete inactivation (as in the case of *Pten/p53* discussed above).

The dogma that every TSG must behave according to the two-hit paradigm could significantly slow the pace of cancer research if potentially critical genes are dismissed as unimportant. The pace of cancer drug development will in turn be delayed by a lack of appreciation of the involvement of these potentially druggable pathways in tumour development and maintenance. A very real challenge for cancer biologists in the next decade will be first to define the TSGs affected by haploinsufficiency and then to model these subtle dosage effects in a way that can yield quantitative models of cancer risk or outcome. Because human genetics analysis alone is often not enough to identify critical haploinsufficient genes unequivocally, mouse models will prove essential to define which genes show haploinsufficiency for TSG function and in which specific cellular and genetic contexts.

Regulation of TSG activity

The continuum model of tumour suppression holds that precise regulation of TSG expression and activity is critical and thus, mechanisms of regulation of expression and function may powerfully contribute to

tumour suppression. One of the most widespread networks of expression regulation is the post-transcriptional regulatory network of miRNAs and competing endogenous RNAs (ceRNAs). In addition to these networks, TSG dosage/activity is regulated by transcriptional, translational and post-translational mechanisms (Fig. 4). These networks open up the realm of genes that could be contextually important for tumorigenesis, especially when considering mechanisms that tune TSG activity and not only those that function as on/off switches.

As an example, a powerful *PTEN*-targeting miRNA network was recently identified. miRNAs are small non-coding RNA molecules that regulate protein expression through inhibition of protein translation or induction of direct cleavage of their mRNA targets via RISC⁵⁷. Many of the *PTEN*-targeting miRNAs are amplified or overexpressed in cancer such as miR-26 in glioma⁵⁸ and miR-22 in prostate cancer⁵⁹. The regulation of coding RNAs by miRNAs is made additionally complex through the interaction of miRNAs with ceRNAs^{60,61}. We coined the term “competing endogenous RNA” (ceRNA) to refer to an RNA—coding or noncoding—that shares miRNA binding sites with another RNA with which it competes for miRNA binding^{60,61}. Any two or more mRNAs with shared miRNA recognition elements (MRE) can act as ceRNAs towards each other by competing for miRNA binding and thereby co-regulating each other's expression. As the expression of the ceRNA increases, the miRNA binding sites on the ceRNA compete with the miRNA binding sites on the other transcript, and less miRNA is available for binding to the TSG transcript (Fig. 4). ceRNA expression (for example, expression of the *PTEN* pseudogene⁶⁰) hence results in sustained TSG expression by preventing miRNA binding to the TSG and subsequent mRNA or protein downregulation. Conversely, loss of ceRNA expression could result in low TSG expression through diversion of additional miRNA molecules to the TSG transcript^{60,61}. Even non-coding ceRNAs that are not expressed at the protein level could function as TSGs by regulating expression of protein-coding TSGs, so mutation or deletion of ceRNA genes could be important in cancer^{60,61}.

Implications for cancer susceptibility and therapy

An appreciation of the continuous nature of tumour suppressive function must in turn lead to a re-evaluation of the role of TSG dosage in cancer susceptibility, diagnosis and treatment. Subtle variation in TSG expression or activity could underlie genetic variation in cancer susceptibility in the population. Inherent, constitutive differences in gene expression could be caused by polymorphic variation in TSG promoter regions or in miRNA binding sites or other regulatory regions. Importantly, synonymous single nucleotide polymorphisms (SNPs) that do not

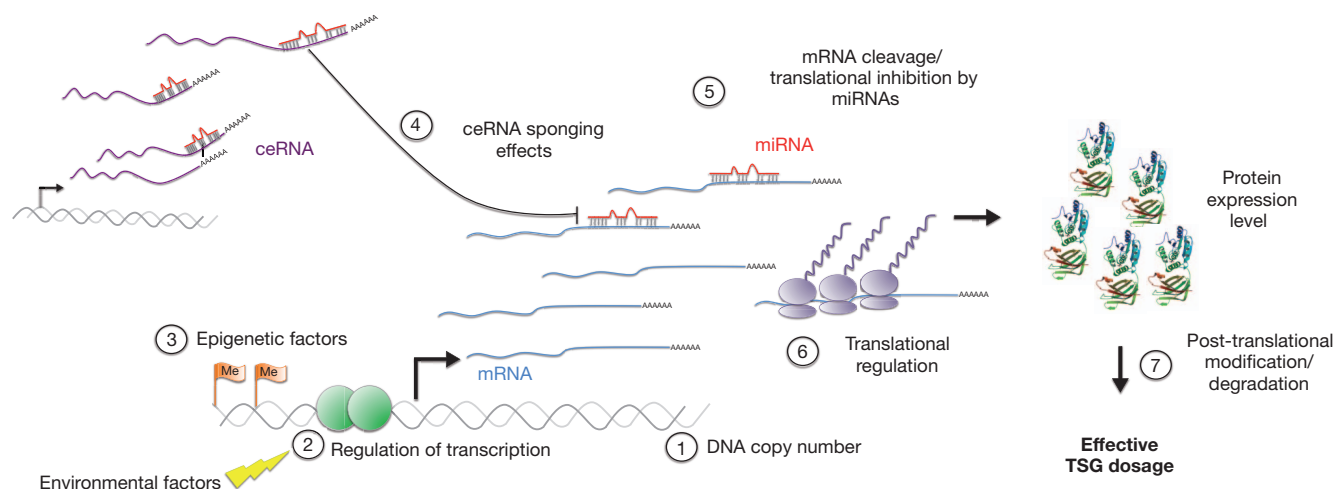


Figure 4 | Mechanisms of regulation of TSG dosage. The interaction between coding and non-coding factors determines final TSG dosage. Classic mechanisms such as DNA copy number (1), transcriptional regulation (2) and epigenetic silencing (3) can affect expression of TSG mRNA. TSG mRNA level or translation into protein is then regulated by miRNAs (5). The availability of

miRNAs for TSG downregulation is further regulated by ceRNA-mediated sponging effects (4). Finally, additional translational regulation (6) or post-translational modifications contribute to the final protein expression, function and effective dosage. The protein structure shown is PTEN (Protein Data Bank accession code 1D5R).

change the protein sequence of an expressed gene could still have a powerful impact by altering the level of TSG expression through alteration of miRNA binding. Moreover, SNPs in non-coding RNAs could have an impact on tumour development through ceRNA effects. Alternatively, on the basis of a continuum model, susceptibility to cancer could be induced by environmental factors through subtle transcriptional up- or downregulation of relevant genes. Development of drugs that prevent and counter these fluctuations could be used as chemopreventative agents to reduce the impact of environmental carcinogens or to correct the effects of inherited predisposing factors.

Just as transcriptional or post-transcriptional regulation may determine cancer susceptibility, it could also be harnessed for cancer prevention and therapy (Fig. 5). As an example, statins, which are already in clinical use, were recently found to upregulate PTEN through PPAR γ signalling⁶². Statins or PTEN-enhancing drugs could therefore be used not only to treat dyslipidaemic cases but also to elevate PTEN expression in patients with tumours with lowered levels of PTEN (Fig. 5). Such an approach would require a quantitative analysis of the expression levels of TSGs and oncogenes and not solely a genetic assessment of their presence or absence.

TSG-targeting microRNAs add another dimension to genotype/phenotype considerations in targeted cancer prevention and therapy. Overexpression of one or more of TSG-targeting miRNAs could confer cancer susceptibility, and tumours not showing genetic alterations in the TSG but with aberrant TSG levels due to miRNA misregulation could behave similarly to tumours with genetic deletion or mutation of the gene. Mapping the interactions between miRNAs and TSGs could be useful for defining and predicting tumour response to therapy.

Provocatively, tumours with partial loss of TSGs like *PTEN* that show obligate haploinsufficiency might be treated with agents to fully suppress the expression of the TSG, thereby suppressing tumour growth via senescence or other fail-safe mechanisms (Fig. 5). Alternatively, when a TSG is completely lost, activation of alternative tumour-suppressive

pathways or synthetic lethality approaches could be considered⁶³ (Fig. 5). The success of such strategies would obviously depend on the integrity of the fail-safe response and the expression level of the TSG, once again requiring a careful annotation of the molecular characteristics of the cancer lesion subjected to treatment. The need for the same careful annotation is not restricted to targeted therapies but applies to any cancer therapy because variations of oncogene or TSG dosage may well be at the core of failure of more conventional radio- or chemotherapies.

On this basis, the scientific challenge will be to develop quantitative assays that can assess and score subtle differences in TSG gene expression to allow accurate diagnostics and prediction of risk/prognosis. Such diagnostics will also be important for analysis and prediction of response to therapeutics, because subtle variations could also determine response to therapy. A tumour with a partial downregulation of a TSG could respond differently from a tumour that has either no alteration of the TSG or complete mutation/loss of the gene (Fig. 5). Critically, in cases where a tumour shows downregulation of a TSG via transcriptional repression, epigenetic downregulation, or aberrant miRNA and ceRNA regulation, there is an opportunity to explore therapeutic restoration of TSG expression.

The recent explosion of knowledge in the field of non-coding RNA biology has opened a new frontier in understanding gene dosage regulation and its involvement in tumorigenesis. In the months and years to come, careful experimentation is required to map out the networks of gene regulation of TSGs and oncogenes to allow drug development and creation of diagnostics from this information. Forty years after the two-hit hypothesis, we are refining and extending our understanding of TSG inactivation in cancer. Milestones made in these past forty years have contributed to the decline in cancer deaths observed in the US since 1990 (ref. 64). The next decades promise further relief from the burden of cancer as we continue to tease out its molecular genetic basis.

Approach	Therapeutic intervention	Applicable tumour type
1 Restoration of TSG pathway or function	- Use drugs that upregulate TSG expression Statins \rightarrow PPAR γ \rightarrow \uparrow PTEN - Inhibition of TSG-downregulating miRNAs miRNA decoys \rightarrow miRNA binding \rightarrow TSG expression - Derepression of epigenetic silencing via DNA demethylating agents or histone modifying drugs DNA methyltransferase inhibitors, histone deacetylase inhibitors	Tumours with partial loss (of TSG expression/activity)
2 Induction/enhancement of fail-safe mechanisms	- Further suppress TSG expression VO-OHpic \rightarrow PTEN Complete loss of PTEN \rightarrow Senescence - Activate alternative TSG pathways Nutlin \rightarrow p53 - Synthetic lethality: use compound genetic effects Sklp2 targeting in PTEN ^{+/+} cells \rightarrow Senescence	Tumours with partial loss Tumours with partial or complete loss

Figure 5 | Opportunities for therapeutic intervention after partial or complete loss of TSG expression. Tumour genotype ultimately determines which therapies can be considered for intervention. Broadly, two general approaches can be taken: restoration of the TSG pathway or function (upper panel) or induction of fail-safe mechanisms by complete downregulation of the TSG (lower panel). In tumours with downregulation of TSG expression in the absence of complete mutation or deletion, expression of the TSG can be attempted using drugs that induce the TSG transcription, inhibit miRNA-induced TSG downregulation, or relieve epigenetic silencing. Alternatively, fail-safe mechanisms may be induced by further downregulation of the TSG, such as inhibition of PTEN by the small molecule VO-OHpic⁶⁹. In tumours with mutation or complete loss of the TSG, activation of other TSG pathways could be pursued. Alternatively, if a synthetic lethality relationship with the TSG is known, then the synthetic lethal gene can be targeted to induce cell death in the cells with complete loss of the TSG.

- Nowell, P. C. & Hungerford, D. A. Chromosome studies on normal and leukemic human leukocytes. *J. Natl Cancer Inst.* **25**, 85–109 (1960).
 - Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
 - Rowley, J. D., Golomb, H. M. & Dougherty, C. 15/17 translocation, a consistent chromosomal change in acute promyelocytic leukaemia. *Lancet* **309**, 549–550 (1977).
 - Heisterkamp, N. *et al.* Localization of the *c-abl* oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. *Nature* **306**, 239–242 (1983).
 - Pandolfi, P. P. *et al.* Structure and origin of the acute promyelocytic leukemia myl/RAR alpha cDNA and characterization of its retinoid-binding and transactivation properties. *Oncogene* **6**, 1285–1292 (1991).
 - Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170–173 (1976).
 - Oppermann, H., Levinson, A. D., Varmus, H. E., Levintow, L. & Bishop, J. M. Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). *Proc. Natl Acad. Sci. USA* **76**, 1804–1808 (1979).
 - Harris, H. The analysis of malignancy by cell fusion: the position in 1988. *Cancer Res.* **48**, 3302–3306 (1988).
 - Knudson, A. G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68**, 820–823 (1971).
- The original statistical analysis of hereditary retinoblastoma that led to the two-hit hypothesis.**
- Lee, W. H. *et al.* Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235**, 1394–1399 (1987).
 - Fung, Y. K. *et al.* Structural evidence for the authenticity of the human retinoblastoma gene. *Science* **236**, 1657–1661 (1987).
 - Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646 (1986).
- Building on several years of work by many laboratories to localize the gene responsible for hereditary retinoblastoma, this study is the first to identify and clone the responsible gene, *RB1*, and to show that it is altered in retinoblastoma tumours.**
- Sparkes, R. S. *et al.* Gene for hereditary retinoblastoma assigned to human chromosome 13 by linkage to esterase D. *Science* **219**, 971–973 (1983).
 - Benedict, W. F. *et al.* Patient with 13 chromosome deletion: evidence that the retinoblastoma gene is a recessive cancer gene. *Science* **219**, 973–975 (1983).
 - Dryja, T. P. *et al.* Homozygosity of chromosome 13 in retinoblastoma. *N. Engl. J. Med.* **310**, 550–553 (1984).
 - Cavenee, W. K. *et al.* Genetic origin of mutations predisposing to retinoblastoma. *Science* **228**, 501–503 (1985).

17. Cavenee, W. K. *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**, 779–784 (1983).
18. Baker, S. J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217–221 (1989).
19. Levy, D. B. *et al.* Inactivation of both APC alleles in human and mouse tumors. *Cancer Res.* **54**, 5953–5958 (1994).
20. Smith, S. A., Easton, D. F., Evans, D. G. & Ponder, B. A. Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genet.* **2**, 128–131 (1992).
21. Gudmundsson, J. *et al.* Different tumor types from BRCA2 carriers show wild-type chromosome deletions on 13q12–q13. *Cancer Res.* **55**, 4830–4832 (1995).
22. Volinia, S. *et al.* Genome wide identification of recessive cancer genes by combinatorial mutation analysis. *PLoS ONE* **3**, e3380 (2008).
23. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
24. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Med.* **10**, 789–799 (2004).
25. Vogelstein, B. & Kinzler, K. W. The multistep nature of cancer. *Trends Genet.* **9**, 138–141 (1993).
26. Hagstrom, S. A. & Dryja, T. P. Mitotic recombination map of 13cen–13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. *Proc. Natl Acad. Sci. USA* **96**, 2952–2957 (1999).
27. Fisher, E. & Scambler, P. Human haploinsufficiency – one for sorrow, two for joy. *Nature Genet.* **7**, 5–7 (1994).
28. Croce, C. M. Oncogenes and cancer. *N. Engl. J. Med.* **358**, 502–511 (2008).
29. Venkatachalam, S. *et al.* Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *EMBO J.* **17**, 4657–4667 (1998).
30. Varley, J. M., Evans, D. G. & Birch, J. M. Li-Fraumeni syndrome – a molecular and clinical review. *Br. J. Cancer* **76**, 1–14 (1997).
31. Clarke, A. R. *et al.* Thymocyte apoptosis induced by p53-dependent and independent pathways. *Nature* **362**, 849–852 (1993).
32. Lynch, C. J. & Milner, J. Loss of one p53 allele results in four-fold reduction of p53 mRNA and protein: a basis for p53 haplo-insufficiency. *Oncogene* **25**, 3463–3470 (2006).
33. Bellacosa, A. *et al.* Altered gene expression in morphologically normal epithelial cells from heterozygous carriers of BRCA1 or BRCA2 mutations. *Cancer Prev. Res.* **3**, 48–61 (2010).
34. Burga, L. N. *et al.* Altered proliferation and differentiation properties of primary mammary epithelial cells from BRCA1 mutation carriers. *Cancer Res.* **69**, 1273–1278 (2009).
35. Proia, T. A. *et al.* Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* **8**, 149–163 (2011).
36. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
37. Sancho, R. *et al.* F-box and WD repeat domain-containing 7 regulates intestinal cell lineage commitment and is a haploinsufficient tumor suppressor. *Gastroenterology* **139**, 929–941 (2010).
38. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
39. Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
40. Lee, J. Y. *et al.* mTOR activation induces tumor suppressors that inhibit leukemogenesis and deplete hematopoietic stem cells after Pten deletion. *Cell Stem Cell* **7**, 593–605 (2010).
41. Kumar, M. S. *et al.* Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* **23**, 2700–2704 (2009).
42. Lambertz, I. *et al.* Monoallelic but not biallelic loss of Dicer1 promotes tumorigenesis in vivo. *Cell Death Differ.* **17**, 633–641 (2010).
43. Bric, A. *et al.* Functional identification of tumor-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16**, 324–335 (2009).
44. Sportoletti, P. *et al.* Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. *Blood* **111**, 3859–3862 (2008).
45. Grisendi, S. *et al.* Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147–153 (2005).
46. Su, X. *et al.* TAp63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature* **467**, 986–990 (2010).
47. Zhou, X. Z. *et al.* The telomerase inhibitor PinX1 is a major haploinsufficient tumor suppressor essential for chromosome stability in mice. *J. Clin. Invest.* **121**, 1266–1282 (2011).
48. Deutschbauer, A. M. *et al.* Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–1925 (2005).
49. Fodde, R. & Smits, R. Cancer biology. A matter of dosage. *Science* **298**, 761–763 (2002).
50. Payne, S. R. & Kemp, C. J. Tumor suppressor genetics. *Carcinogenesis* **26**, 2031–2045 (2005).
51. Trotman, L. C. *et al.* Pten dose dictates cancer progression in the prostate. *PLoS Biol.* **1**, E59 (2003).
52. Alimonti, A. *et al.* Subtle variations in Pten dose determine cancer susceptibility. *Nature Genet.* **42**, 454–458 (2010).
53. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
54. Evan, G. I. *et al.* Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69**, 119–128 (1992).
55. Takano, T. *et al.* Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J. Clin. Oncol.* **23**, 6829–6837 (2005).
56. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
57. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
58. Huse, J. T. *et al.* The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis *in vivo*. *Genes Dev.* **23**, 1327–1337 (2009).
59. Poliseno, L. *et al.* Identification of the miR-106b~25 microRNA cluster as a proto-oncogenic PTEN-targeting intron that cooperates with its host gene MCM7 in transformation. *Sci. Signal.* **3**, ra29 (2010).
60. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
61. Salmerna, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* doi:10.1016/j.cell.2011.07.014 (2011).
62. Teresi, R. E., Planchon, S. M., Waite, K. A. & Eng, C. Regulation of the PTEN promoter by statins and SREBP. *Hum. Mol. Genet.* **17**, 919–928 (2008).
63. Lin, H. K. *et al.* Skp2 targeting suppresses tumorigenesis by Arf-p53-independent cellular senescence. *Nature* **464**, 374–379 (2008).
64. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J. Clin.* **60**, 277–300 (2010).
65. Di Cristofano, A. *et al.* Impaired Fas response and autoimmunity in Pten^{+/-} mice. *Science* **285**, 2122–2125 (1999).
66. Suzuki, A. *et al.* High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Curr. Biol.* **8**, 1169–1178 (1998).
67. Podsypanina, K. *et al.* Mutation of Pten/Mmac1 in mice causes neoplasia in multiple organ systems. *Proc. Natl Acad. Sci. USA* **96**, 1563–1568 (1999).
68. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. Pten is essential for embryonic development and tumour suppression. *Nature Genet.* **19**, 348–355 (1998).
69. Alimonti, A. *et al.* A novel type of cellular senescence that can be enhanced in mouse models and human tumor xenografts to suppress prostate tumorigenesis. *J. Clin. Invest.* **120**, 681–693 (2010).
70. Daikoku, T. *et al.* Conditional loss of uterine Pten unfaithfully and rapidly induces endometrial cancer in mice. *Cancer Res.* **68**, 5619–5627 (2008).
71. Li, G. *et al.* Conditional loss of PTEN leads to precocious development and neoplasia in the mammary gland. *Development* **129**, 4159–4170 (2002).

Acknowledgements We thank L. Salmerna, L. Poliseno and all Pandolfi laboratory members for advice and critical discussions. This work was supported in part by NIH core grant CA06927 and an appropriation from the Commonwealth of Pennsylvania to the Fox Chase Cancer Center to A.G.K. and NIH grant R01CA142787 to A.H.B. and P.P.P.

Author Contributions A.H.B., A.G.K. and P.P.P. together contributed to all aspects of this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to P.P.P. (ppandolfi@bidmc.harvard.edu).

The landscape of recombination in African Americans

A list of authors and their affiliations appears at the end of the paper

Recombination, together with mutation, gives rise to genetic variation in populations. Here we leverage the recent mixture of people of African and European ancestry in the Americas to build a genetic map measuring the probability of crossing over at each position in the genome, based on about 2.1 million crossovers in 30,000 unrelated African Americans. At intervals of more than three megabases it is nearly identical to a map built in Europeans. At finer scales it differs significantly, and we identify about 2,500 recombination hotspots that are active in people of West African ancestry but nearly inactive in Europeans. The probability of a crossover at these hotspots is almost fully controlled by the alleles an individual carries at *PRDM9* (P value $< 10^{-245}$). We identify a 17-base-pair DNA sequence motif that is enriched in these hotspots, and is an excellent match to the predicted binding target of *PRDM9* alleles common in West Africans and rare in Europeans. Sites of this motif are predicted to be risk loci for disease-causing genomic rearrangements in individuals carrying these alleles. More generally, this map provides a resource for research in human genetic variation and evolution.

In humans and many other species, recombination is not evenly distributed across the genome, but instead occurs in 'hotspots': 2-kilobase (kb) segments where the crossover rate is far higher than in the flanking DNA sequence^{1–3}. The highest-resolution genetic map in contemporary humans so far—the deCODE map—is based on about 500,000 crossovers identified in 15,000 Icelandic meioses⁴. However, a limitation of maps built in people of European descent^{4–6} is that they may not apply equally well in other populations, as suggested by comparisons of maps across ethnic groups^{4,7–9} and patterns of linkage disequilibrium breakdown, which indicate that more of the genome may be recombinationally active in West Africans¹⁰. It is known that a major determinant of the positions of recombination hotspots is *PRDM9*, a meiosis-specific histone H3 methyltransferase whose zinc finger (ZF) domain binds DNA sequence motifs^{11–13}. In Europeans, *PRDM9* ZF arrays are predominantly of two similar types, A and B, both of which bind the 13-bp motif CCNCCNTNCCNC¹¹. In contrast, 36% of West African alleles are not of the A or B type^{9,13}. Sperm typing of males who carry neither the A nor the B allele has shown no evidence of crossover activity at recombination hotspots associated with the 13-bp motif⁹.

Building an African–American genetic map

To investigate differences in the crossover landscape across human populations, we built a genetic map in African Americans, who have an average of about 80% West African and 20% European ancestry, leading to genomes comprised of multi-megabase stretches of either West African or European ancestry¹⁴. Computational approaches, including HAPMIX¹⁵, have been developed to infer the probability of 0, 1 or 2 European or African alleles at each locus in individuals genotyped at hundreds of thousands of single nucleotide polymorphisms (SNPs)^{15–17}. Positions where the inferred number of European or African alleles changes reflect crossover events that have occurred since admixture began (on average six generations ago¹⁵). Change in the probability of European ancestry between adjacent SNPs can be interpreted as the probability of such a crossover between them. We inferred crossover events in 29,589 apparently unrelated African Americans who had been genotyped on SNP arrays in genetic association studies (Methods; Fig. 1a). To minimize false-positive crossovers, we restricted

to crossovers that HAPMIX inferred with a probability of $>95\%$, and that were flanked by a minimum of 2-centimorgan (cM) stretches where the ancestry was inferred to be unchanging (Supplementary Note 1). This produced 2,113,293 high-confidence crossovers, with a typical switch point resolved within 70 kb with probability 50% (Supplementary Note 1).

To build a high-resolution African–American genetic map (AA map), we leveraged the fact that most crossovers occur in hotspots shared across individuals² (Methods). Intuitively, although any crossover can only be roughly localized, inter-SNP intervals that are inferred to have an appreciable probability of crossover in multiple individuals are likely to contain recombination hotspots, allowing much better localization (Supplementary Fig. 1). To implement this idea, we modelled the recombination rate for each inter-SNP interval as shared across individuals and used Markov chain Monte Carlo (MCMC) to sample rates consistent with the data (Methods). This provides well-calibrated estimates of the crossing-over rate between all pairs of markers as well as estimates of rate uncertainty (Supplementary Note 1 and Supplementary Fig. 2). We find that the interval size at which the average recombination rate is equal to the standard error is 6 kb, which is the same accuracy that would be expected from a map based on 500,000 crossovers whose boundaries were precisely resolved (Supplementary Note 1). Despite this high resolution, there are also some limitations. First, the AA map does not separately infer male and female recombination rates (it is a sex-averaged map) and requires normalization by the total map length (like linkage disequilibrium maps^{3,18}). Second, the map has less resolution and may miss a higher fraction of true crossovers at loci where it is more difficult to detect and resolve crossovers owing to low SNP density or low differentiation between West Africans and Europeans. Third, the map may be biased where ancestry deviates from the average, for example at chromosome 8q24, where the 10% of the people in this study who have prostate cancer have an increased proportion of African ancestry¹⁹. Fourth, the map assumes that all individuals are unrelated, whereas in fact there is probably some shared ancestry, resulting in multiple counting of some crossovers and an overestimation of map precision.

To assess the accuracy of the AA map, we generated an independent African–American pedigree map by analysing 222 nuclear families

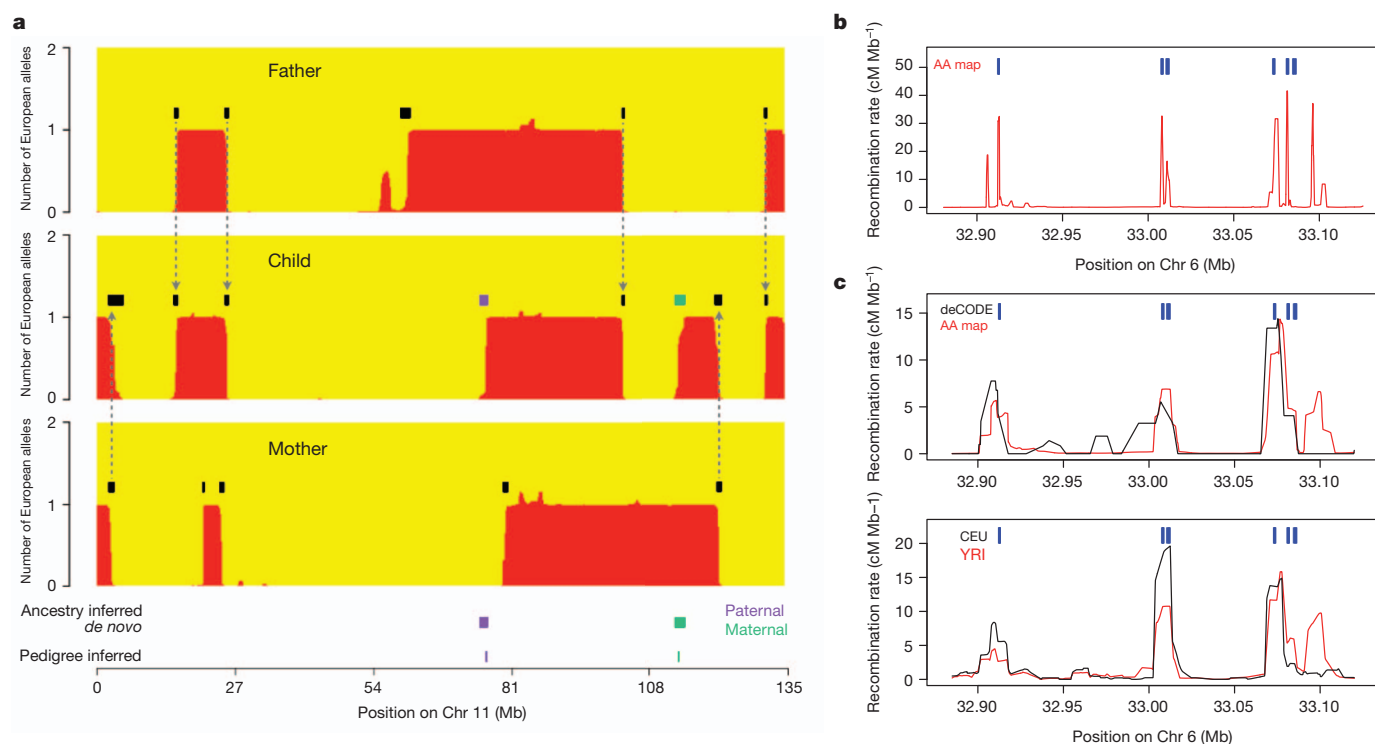


Figure 1 | Building an African-American genetic map. **a**, HAPMIX detection of crossovers between segments of inferred ancestry is illustrated in a father-mother-child trio. Black segments show inferred crossovers; arrows show transmission of ancestral crossovers from parent to child; purple/green segments show *de novo* events (paternal/maternal origin, respectively) corresponding to

that included 1,056 meioses in which we could directly detect crossovers between parent and child (Methods; Fig. 1a). Examination of the AA map rate around directly detected crossovers confirms the high resolution: the rate around such crossovers shows at least as strong a peak as that observed in maps based on linkage disequilibrium^{2,3,18} (Supplementary Fig. 3). We next computed correlation coefficients for both the AA map and the deCODE map⁴ to maps derived from the breakdown of linkage disequilibrium in Europeans (CEU) and West Africans (YRI)¹⁸. At broad scales (>3 Mb) they are almost identical ($\rho > 0.97$; Table 1). At fine scales, the AA map is more accurate (Table 1 and Supplementary Table 1), as reflected in a modest improvement in correlation to the CEU map at a 3-kb scale ($\rho_{AA,CEU} = 0.66$ versus $\rho_{deCODE,CEU} = 0.58$), and a major improvement for the YRI map, also at a 3-kb scale ($\rho_{AA,YRI} = 0.71$ versus $\rho_{deCODE,YRI} = 0.53$). The deCODE map is more correlated to the CEU map than to the YRI map at scales <1 Mb, suggesting that this map, built in Icelanders, reflects more European recombination rates. The AA map shows the opposite pattern, suggesting that it reflects more West African recombination patterns.

events identified directly using two additional children (bottom, 'pedigree inferred'). **b**, The AA map localizes five hotspots in a region of the MHC whose positions (blue) were previously mapped by sperm typing¹. **c**, Comparison of maps shows a hotspot at 33.1 Mb in the African-derived AA and YRI maps, but not the deCODE and CEU maps (all maps smoothed to 10 kb).

Population differences in hotspot locations

We compared the rate estimates for all four maps (AA, deCODE, CEU and YRI) over a 200-kb region within the major histocompatibility complex (MHC) locus where recombination rates in European males have been characterized through sperm typing¹ (Fig. 1b). The AA map detects five of six known hotspots, and localizes them to within 1 kb (the sixth hotspot is weak, with a peak male rate below the genome average¹). Notably, the two maps based on samples with African ancestry (AA and YRI) found a hotspot not present in either map based on samples of European ancestry (deCODE and CEU) (Fig. 1c; Supplementary Fig. 4 gives a second example). We confirmed that such 'African-enriched' hotspots also occur genome-wide, by examining 2,375 loci with recombination rate peaks in the YRI map ($>5 \text{ cM Mb}^{-1}$) but not the CEU map ($<1 \text{ cM Mb}^{-1}$), and finding a rate rise in the independently generated AA map, but not in the deCODE map (Supplementary Fig. 5A). In the reciprocal experiment searching for European-specific hotspots, we find no such evidence for genuine ancestry specificity; at loci with recombination rate peaks in the CEU map but not the YRI map, there are weak peaks in both the deCODE and AA maps

Table 1 | Genetic map assessments at different size scales

Scale (interval size)	Pearson correlation (ρ) of the AA map (deCODE map) to the specified LD map			Estimated correlation of AA map to the true map (inferred by MCMC) [†]	Estimated coefficient of variation of AA map (s.e. divided by crossover rate expected for interval size) [†]
	Combined LD*	CEU	YRI		
3 kb	0.75 (0.63)	0.66 (0.58)	0.71 (0.53)	0.93	1.41
10 kb	0.82 (0.74)	0.73 (0.70)	0.78 (0.65)	0.96	0.73
30 kb	0.86 (0.83)	0.78 (0.78)	0.83 (0.74)	0.98	0.36
100 kb	0.91 (0.89)	0.84 (0.85)	0.87 (0.81)	0.99	0.17
300 kb	0.94 (0.93)	0.89 (0.90)	0.92 (0.88)	1.00	0.08
1 Mb	0.97 (0.96)	0.94 (0.94)	0.95 (0.95)	1.00	0.04
3 Mb	0.98 (0.98)	0.97 (0.97)	0.98 (0.97)	1.00	0.02

The numbers in this table are restricted to the autosomes and genomic segments more than 5 Mb from the telomeres. LD, linkage disequilibrium; s.e., standard error.

* The combined map is the HapMap2 population-averaged linkage-disequilibrium-based map¹⁸.

[†] The s.e. of the map at each size scale is determined by the posterior probability distribution from the MCMC.

(Methods and Supplementary Fig. 5B). Thus, hotspots active in Europeans are consistently ‘shared’ with YRI and African Americans, whereas populations with African ancestry harbour additional, non-shared hotspots that we call ‘African-enriched’.

Mapping variants underlying population differences

To understand the features of recombination in West Africans that differ from Europeans, we estimated the degree to which each African-American person’s crossovers occur in African-enriched hotspots, compared with shared hotspots, a phenotype we refer to as their African enrichment (AE). We view each individual’s crossovers as sampled from a mixture of two genetic maps—an ‘S map’ of shared hotspots based on the deCODE map, and an ‘AE map’ of African-enriched hotspots that is learned from comparing the deCODE and AA maps—so that the proportion of crossovers assigned to the AE map is a person’s AE phenotype (Supplementary Note 4). We tested approximately 3 million SNPs (genotyped and imputed) for association with three phenotypes: AE, usage of linkage-disequilibrium-based hotspots known to be enriched for the 13-bp motif

CCNCCNTNNCCNC²⁰ and genome-wide crossover rate (in pedigrees) (Methods and Supplementary Note 4). In crossovers detected in unrelated African Americans, the alleles a person carries are only sometimes descended from the ancestor in whom the crossover occurred, thus adding noise to the association signal (nevertheless there is useful signal given the large sample size; Supplementary Note 4). In the pedigree map, association between alleles and AE can be tested directly because we have genotypes in the parents.

The SNP showing the strongest association with AE is rs6889665 ($P = 1.5 \times 10^{-246}$; Fig. 2a and Supplementary Fig. 6), which has a derived allele frequency of 29% in YRI and 2% in CEU, and is within 4 kb of the ZF array of *PRDM9* (refs 4, 9, 11–13). This SNP is associated with AE in both the pedigree individuals and the unrelated individuals (Supplementary Note 4), and is also the SNP most strongly associated with usage of linkage-disequilibrium-based hotspots ($P = 1.8 \times 10^{-52}$) (Supplementary Table 2). No locus outside *PRDM9* is significant ($P < 0.01$ after Bonferroni correction; Supplementary Table 2). To understand better the association at rs6889665, we inferred the alleles in the *PRDM9* ZF array carried

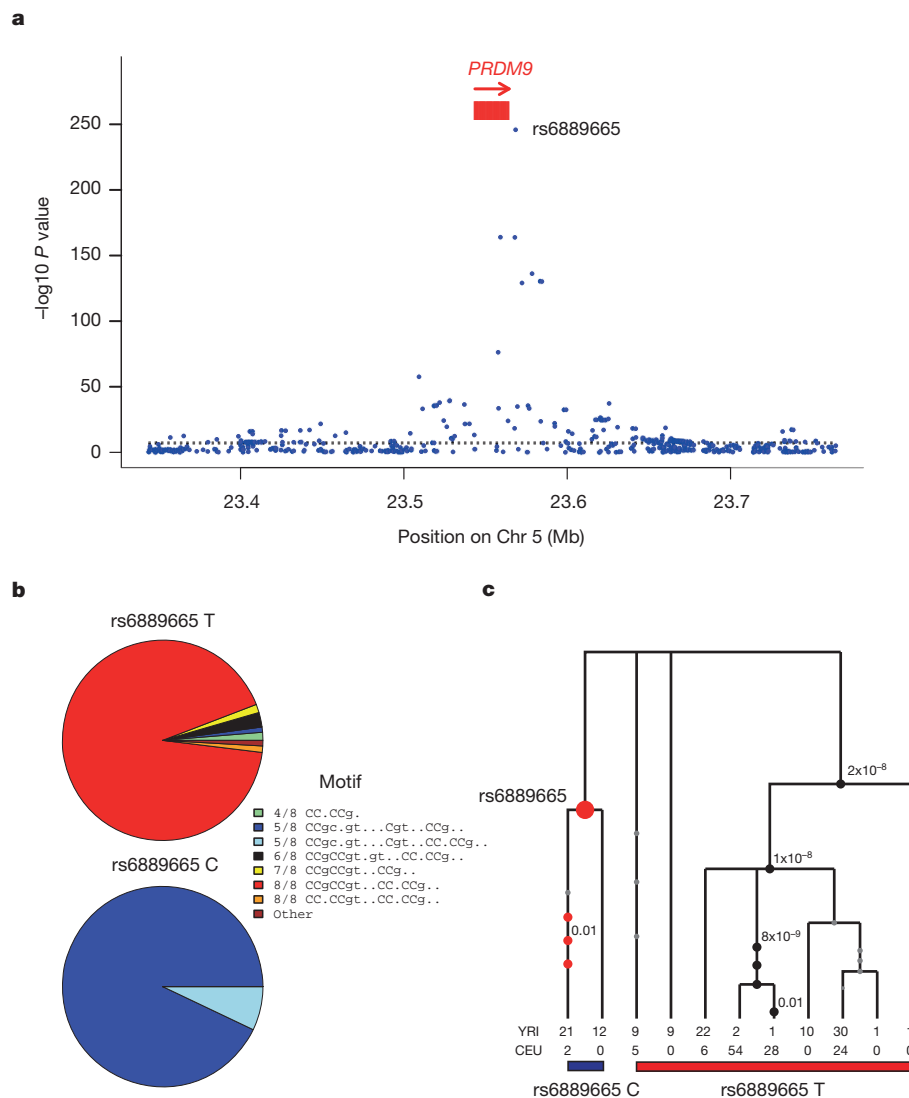


Figure 2 | Association of *PRDM9* genetic variation with hotspot activity. **a**, A genome-wide association study measuring association of the AE phenotype shows a single genome-wide significant peak at *PRDM9*, with rs6889665 the best-associated SNP. **b**, Relationship between alleles of rs6889665 and predicted binding target of the *PRDM9* ZF array⁹ for West African and European samples. The binding predictions are grouped into 8

clusters according to their best-matching region to the 13-bp motif, and annotated by the number of bases matching the motif. The African-enriched rs6889665 C allele always co-occurs with motifs with a poor (5/8) match to the 13-bp motif. **c**, Gene tree²⁵ of the linkage disequilibrium block containing the *PRDM9* ZF array (Methods); numbered circles show SNPs and significant P values for association, after conditioning on rs6889665.

by 139 individuals based on sequencing data from the 1000 Genomes Project¹⁰, using the reads to infer each individual's *PRDM9* alleles among 29 alleles whose full sequences were previously determined⁹ (Supplementary Note 5). Grouping *PRDM9* alleles on the basis of how closely their binding target predictions match the 8 non-degenerate bases of the 13-bp motif, following a previously described approach⁹, we find that the ancestral 'T' variant at rs6889665 is strongly correlated to alleles with an exact (8/8) match to the 13-bp motif (including the A and B alleles), whereas the derived 'C' variant is almost perfectly correlated to a group of alleles, all predicted to bind a common, different 17-bp motif—CCgCNgTNNCgtNNCC⁹—which matches the 13-bp motif at only 5 bases (5/8 match; less strongly signalled bases in the motif are in lowercase and 'N' may be any base). This implies a common historical origin for alleles matching this 17-bp motif (Fig. 2b, Supplementary Fig. 7 and Supplementary Note 5). We also experimentally measured the number of ZF domains in *PRDM9* in 354 individuals including 166 African Americans from the pedigree study (Methods). This showed, again, that rs6889665 differentiates *PRDM9* alleles into two different classes, with 96% of haplotypes carrying the ancestral allele having <14 ZFs, and 93% of haplotypes carrying the derived allele having ≥14 ZFs (Supplementary Fig. 7). After conditioning on rs6889665, there is no evidence that ZF array length is associated with the AE phenotype. Several SNPs near the *PRDM9* ZF array show a conditional association signal that is much weaker than rs6889665, but still significant (Fig. 2c, Supplementary Fig. 6 and Supplementary Note 4), with the strongest at rs10043097 ($P = 8.3 \times 10^{-14}$), upstream of the *PRDM9* transcription start site. These SNPs may tag additional variation in the *PRDM9* ZF array, or potentially expression levels.

Finding a motif for African-enriched hotspots

To identify directly candidate African-enriched hotspot motifs, we selected 2,454 loci with a high crossover rate in the AE map and YRI map (>2 cM Mb⁻¹ over 2 kb), and no more than half this rate in the S map and CEU map (this set is more powerfully enriched for higher recombination in people of African ancestry than the 2,375 above, as it includes information from the contemporary maps). We compared these to a 'control set' of 7,328 candidate hotspots more active in the European- than the African-derived maps (Methods and Supplementary Note 6). To identify sequence motifs associated with the African-enriched hotspots^{3,21}, we identified short motifs that

occurred at increased frequency in the African-enriched hotspot set (Supplementary Note 6). Testing all motifs with lengths of 5–9 bases revealed a 9-nucleotide motif CCCCAGTGA (odds ratio (OR) = 1.79, $P = 2.24 \times 10^{-8}$, Bonferroni corrected $P = 0.004$), which exhibited a kilobase-scale rate peak near occurrences of this motif in African-derived maps, but in neither of the European-derived maps (Supplementary Fig. 8). Further analysis revealed a strong influence of downstream flanking bases (Supplementary Fig. 9) and degeneracy, yielding a 17-bp consensus sequence, CCCCAGTGAGCGTtGcC (Fig. 3a; more strongly signalled bases are in uppercase), with the same consensus obtained when we considered flanking sequences for only odd or even chromosomes, and whether we based the analysis on AE-S or YRI-CEU map comparisons (Supplementary Note 6). The 500 best matches to this motif have a ~3-fold increase in average rate in the AA and YRI relative to the deCODE and CEU maps (Fig. 3b and Supplementary Fig. 8G). Hotspots associated with the motif occur in both unique and repetitive DNA (for example, LIPA10/13 LINE elements; Supplementary Fig. 10 and Supplementary Note 6). We also compared the 17-bp consensus to the binding motif predicted for 5/8 match alleles, and found that they match almost precisely (Fig. 3a; 10 of 11 bases, $P = 8.1 \times 10^{-6}$).

Assessing the impact of *PRDM9* on recombination

How much of the African-enriched recombination pattern can be explained by *PRDM9*? We estimated the fraction of variation in the AE phenotype explained by rs6889665 in our pedigree data after accounting for noise in the phenotype estimation (Supplementary Note 4). Over 82% of map usage variability is explained by the rs6889665 genotype alone. Given that there are further influential *PRDM9* variants (Fig. 2c), this gene may thus explain almost all differences in local rate between the West African and European populations. We next examined rates around 82 narrowly defined (<10 kb) crossover sites in 7 individuals homozygous for the derived allele at rs6889665. There is no evidence of hotspots at these loci in either the deCODE or CEU maps (Fig. 3c), in contrast to crossovers in individuals carrying the ancestral allele at rs6889665 (Supplementary Fig. 11). Thus, crossover positions in individuals who are homozygous for the derived allele at rs6889665 are consistent with an entirely different recombination hotspot landscape, which would imply *PRDM9* control of all hotspots⁹. Despite the strong correlation between maps at megabase scales, there is mounting evidence that *PRDM9*'s influence

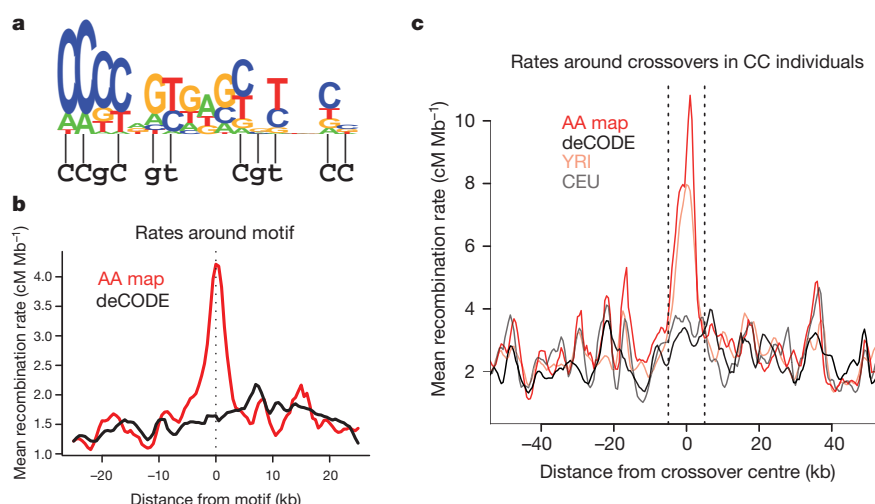


Figure 3 | A sequence motif specifying the positions of African-enriched hotspots. **a**, Logo plot showing a degenerate 17-bp hotspot motif, with stack height proportional to $-\log P$ value, and relative letter height proportional to the mean crossover rate increase given each base. Below is the bioinformatic *PRDM9* binding prediction for the alleles associated with rs6889665 allele C (from Fig. 2b), matching this motif at 10/11 bases (lines). **b**, Average crossover

rate (in 2-kb sliding windows) in the AA (red line) and deCODE (black line) maps surrounding the 500 strongest motif matches. **c**, In seven rs6889665 CC individuals from the pedigree study, we localized 82 crossovers to within 10 kb, and plot average AA, YRI, deCODE and CEU map rates. There is no strong peak above local background in the deCODE or CEU maps.

on crossing over may not be limited to fine scales^{4,11}: we observe a weakly significant association of rs6889665 with the total number of crossovers genome-wide in pedigrees ($P = 0.04$), corresponding to an average 1.3 crossovers more per meiosis per derived allele, exceeding the strongest previously known association²² at *RNF212*.

Conclusions

We have shown that *PRDM9* alleles that bind a novel 17-bp motif and occur at greatly increased frequency in people of West African ancestry have led to a shift in the recombination landscape compared with people of non-African ancestry. The larger number of hotspots available to West Africans implies that at the population level, crossovers are more evenly distributed than in Europeans¹⁰, and thus the shorter extent of West African linkage disequilibrium is not due to differences in demographic history alone (such as the lack of an out-of-Africa founder event)²³. Our findings also have medical implications, as recombination errors leading to insertions or deletions are known to be associated with recombination hotspots^{9,21,24}. Our results predict that the congenital abnormalities that have been associated with the recombination hotspots bound by *PRDM9* A and B alleles will occur at a decreased rate in people of West African ancestry, whereas new diseases will arise due to recombination errors near African-enriched hotspots.

METHODS SUMMARY

We assembled SNP array data from 29,589 unrelated people and 222 nuclear families genotyped at 490,000–910,000 SNPs from the Candidate Gene Association Resource (CARE), studies at the Children's Hospital of Philadelphia (CHOP), the African American Breast Cancer Consortium, the African American Prostate Cancer Consortium and the African American Lung Cancer Consortium. To build a recombination map, we used HAPMIX to localize candidate crossover positions¹⁵, and implemented a MCMC that used the probability distributions for the positions of the filtered crossovers to infer recombination rates for each of 1.3 million inter-SNP intervals. We also implemented a second MCMC that models each individual's set of crossovers as a mixture of an S map, similar to the European deCODE map, and an AE map, and then assigned each individual an 'AE phenotype' corresponding to the proportion of their newly detected crossovers assigned to the AE map. We imputed genotypes at up to three million HapMap2 SNPs¹⁸ and then tested each of these SNPs for association with the AE phenotype and other recombination-related phenotypes. We identified 2,454 candidate African-enriched hotspots with increased recombination rates in the YRI versus CEU maps, and in the AE versus S maps, and searched for motifs enriched at these loci, thus identifying a degenerate 17-bp motif. To study the structure of *PRDM9*, we measured the length of the *PRDM9* ZF array and genotyped rs6889665 in YRI, CEU and the CARE nuclear families; we also carried out imputation based on 1000 Genomes Project short read data¹⁰ to infer the alleles individuals carry, among 29 previously characterized in a sequencing study of *PRDM9* (ref. 9).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 February; accepted 27 June 2011.

Published online 20 July 2011.

1. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
2. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
3. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
4. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
5. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
6. Matise, T. C. *et al.* A second-generation combined linkage-physical map of the human genome. *Genome Res.* **17**, 1783–1786 (2007).
7. Weitkamp, L. R. Proceedings: population differences in meiotic recombination frequency between loci on chromosome 1. *Cytogenet. Cell Genet.* **13**, 179–182 (1974).
8. Jorgenson, E. *et al.* Ethnicity and human genetic linkage maps. *Am. J. Hum. Genet.* **76**, 276–290 (2005).

9. Berg, I. L. *et al.* *PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genet.* **42**, 859–863 (2010).
10. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
11. Baudat, F. *et al.* *PRDM9* is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
12. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).
13. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
14. Smith, M. W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74**, 1001–1013 (2004).
15. Price, A. L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
16. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* **82**, 290–303 (2008).
17. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000 (2004).
18. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
19. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103**, 14068–14073 (2006).
20. Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398 (2008).
21. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hotspots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).
22. Kong, A. *et al.* Sequence variants in the *RNF212* gene associate with genome-wide recombination rate. *Science* **319**, 1398–1401 (2008).
23. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
24. Raedt, T. D. *et al.* Conservation of hotspots for recombination in low-copy repeats associated with the *NF1* microdeletion. *Nature Genet.* **38**, 1419–1423 (2006).
25. Griffiths, R. C. & Tavaré, S. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.* **127**, 77–98 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the participants who donated DNA samples, to D. Altshuler, J. Buard, K. Bryc, J. Kovacs, B. de Massy, G. McVean, B. Pasaniuc and S. Sankararaman for conversations and critiques, and to A. Auton for facilitating analysis of the 1000 Genomes Project data. Analysis was supported by the Wellcome Trust and NIH grants HL084107 and GM091332. CARE was supported by a contract from the National Heart, Lung and Blood Institute (HHSN268200960009C) to create a phenotype and genotype database for dissemination to the biomedical research community. Eight parent studies contributed phenotypic data and DNA samples through the Broad Institute (N01-HC-65226): the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA) study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS) and the Sleep Heart Health Study (SHHS). Support for CARE also came from the individual research institutions, investigators, field staff and study participants. Individual funding information is available at <http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx>. All genome-wide genotyping of samples from the Children's Hospital of Pennsylvania (CHOP) was supported by an Institutional Development Award to the Center for Applied Genomics from the Children's Hospital of Philadelphia, a research award from the Landenberger Foundation and the Cotswold Foundation. We thank all study participants and the staff at the Center for Applied Genomics for performing the genotyping. The African American Breast Cancer Consortium (AABCC) was supported by a DoD Breast Cancer Research Program Era of Hope Scholar Award to C.A.H. and the Norris Foundation, and by grants to the component studies: MEC (CA63464, CA54281); CARE (HD33175); WCHS (CA100598, DAMD 170100334, Breast Cancer Research Foundation); SFBC (CA77305, DAMD 17966071); CBCS (CA58223, ES10126), PLCO (NCI Intramural Research Program); NHBS (CA100374); WFBC (R01-CA73629); and CPS-II (the American Cancer Society). The African American Prostate Cancer Consortium (AAPCC) was supported by grants CA63464, CA54281, CA1326792, CA148085 and HG004726, and by grants to the component studies: PLCO (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), both MEC and LAAPC (PC35139, DP000807); MDA (CA68578, CA140388, ES007784, DAMD W81XWH0710645); GECAP (ES011126); CaP Genes (CA88164); IPCG (W81XWH0710122); DCPC (GM08016, DAMD W81XWH0710203, DAMD W81XWH0610066); and SCCS (CA092447, CA68485). The African American Lung Cancer Consortium (AALCC) was supported by grants CA060691, CA87895, PC35145 and CA22453, CA68578, CA140388, ES007784, ES06717, CA55769, CA127219, CA1116460S1, CA1116460, CA121197, CA141716, CA121197S2, CPRIT RP100443, CA148127, DAMD W81XWH0710645, University Cancer Foundation, Duncan Family Institute, Center for Community, Implementation and Dissemination Research Core, and by grants to the component studies: PLCO and the Maryland Studies (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), and both MEC and LAAPC (PC35139, DP000807).

Author Contributions D.R. and S.R.M. conceived the study. A.G.H., A.T., N.P., Y.S., N.R., C.D.P., G.K.C., K.W., S.G.B., D.R. and S.R.M. performed analyses. N.R. performed the experimental work (genotyping of polymorphisms at *PRDM9*). A.G.H., N.P., J.N.H.,

B.E.H., H.A.T. Jr, A.L.P., H.H., S.J.C., C.A.H., J.G.W., D.R. and S.R.M. coordinated the study. A.G.H., D.R. and S.R.M. wrote the paper. N.R., C.D.P., G.K.C., K.W., S.G.B., S.R., J.N.H., B.E.H., H.A.T. Jr, H.H., S.J.C., C.A.H., J.G.W., D.R. and all the alphabetically listed authors contributed to sample collection and generation of SNP array data. All authors contributed to revision and review of the manuscript.

Author Information Crossover rate estimates for the AA map can be found at <http://www.well.ox.ac.uk/~anjali/AAmap/>. We also provide estimates of uncertainty for the map based on samples from the MCMC. Association testing results for each SNP are available from the authors on request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or S.R.M. (myers@stats.ox.ac.uk).

Anjali G. Hinch¹, Arti Tandon^{2,3}, Nick Patterson², Yunli Song⁴, Nadin Rohland^{2,3}, Cameron D. Palmer^{5,6}, Gary K. Chen⁷, Kai Wang^{8,9}, Sarah G. Buxbaum¹⁰, Ermeg L. Akylbekova^{10,11}, Melinda C. Aldrich^{12,13}, Christine B. Ambrosone¹⁴, Christopher Amos¹⁵, Elisa V. Bandera¹⁶, Sonja I. Berndt¹⁷, Leslie Bernstein¹⁸, William J. Blot^{13,19}, Cathryn H. Bock²⁰, Eric Boerwinkle²¹, Qiuyin Cai¹³, Neil Caporaso¹⁷, Graham Casey⁷, L. Adrienne Cupples²², Sandra L. Deming¹³, W. Ryan Diver²³, Jasmin Divers²⁴, Myriam Fornage²⁵, Elizabeth M. Gillanders²⁶, Joseph Glessner⁹, Curtis C. Harris²⁷, Jennifer J. Hu²⁸, Sue A. Ingles⁷, William Isaacs²⁹, Esther M. John³⁰, W. H. Linda Kao³¹, Brendan Keating⁹, Rick A. Kittles³², Laurence N. Kolonel³³, Emma Larkin³⁴, Loic Le Marchand³³, Lorna H. McNeill³⁵, Robert C. Millikan³⁶, Adam Murphy³⁷, Solomon Musani¹¹, Christine Neslund-Dudas³⁸, Sarah Nyante³⁶, George J. Papanicolaou³⁹, Michael F. Press⁷, Bruce M. Psaty⁴⁰, Alex P. Reiner⁴¹, Stephen S. Rich⁴², Jorge L. Rodriguez-Gil²⁸, Jerome I. Rotter⁴³, Benjamin A. Rybicki³⁸, Ann G. Schwartz²⁰, Lisa B. Signorello^{13,19}, Margaret Spitz¹⁵, Sara S. Strom⁴⁴, Michael J. Thun²³, Margaret A. Tucker¹⁷, Zhaoming Wang⁴⁵, John K. Wiencke⁴⁶, John S. Witte⁴⁷, Margaret Wrensch⁴⁶, Xifeng Wu¹⁵, Yuko Yamamura⁴⁴, Krista A. Zanetti^{26,27}, Wei Zheng¹³, Regina G. Ziegler¹⁷, Xiaofeng Zhu⁴⁸, Susan Redline⁴⁹, Joel N. Hirschhorn^{5,6,50}, Brian E. Henderson⁷, Herman A. Taylor Jr^{11,51,52}, Alkes L. Price⁵³, Hakon Hakonarson^{9,54}, Stephen J. Chanock¹⁷, Christopher A. Haiman⁷, James G. Wilson⁵⁵, David Reich^{2,3,*} & Simon R. Myers^{1,4,*}

¹Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK. ²Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³Department of Genetics, Harvard Medical School, New Research Building, 77 Ave. Louis Pasteur, Boston, Massachusetts 02115, USA.

⁴Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK.

⁵Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁶Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital Boston, Massachusetts 02115, USA.

⁷Department of Preventive Medicine and Department of Pathology, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California 90033, USA. ⁸Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, California 90089, USA. ⁹Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ¹⁰Jackson Heart Study Coordinating Center, Jackson State University, 350 W. Woodrow Wilson Ave., Suite 701, Jackson, Mississippi 39213, USA. ¹¹Department of Medicine, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA.

¹²Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. ¹³Division of Epidemiology in the Department of Medicine, Vanderbilt Epidemiology Center; and the Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. ¹⁴Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York 14263, USA.

¹⁵Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ¹⁶The Cancer Institute of New Jersey, New Brunswick, New Jersey 08903, USA. ¹⁷Division of

Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. ¹⁸Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, California 91010, USA. ¹⁹International Epidemiology Institute, Rockville, Maryland 20850, USA. ²⁰Karmanos Cancer Institute and Department of Oncology, Wayne State University of Medicine, Detroit, Michigan 48201, USA. ²¹Human Genetics Center and Division of Epidemiology, University of Texas at Houston, 1200 Herman Pressler St., Houston, Texas 77030, USA. ²²Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, Massachusetts 02118 and Framingham Heart Study, Framingham, Massachusetts 01702, USA.

²³Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. ²⁴Department of Biostatistical Sciences, Wake Forest University School of Medicine WC-2326, Medical Center Blvd., Winston Salem, North Carolina 27157, USA. ²⁵Institute of Molecular Medicine and Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center at Houston, 1825 Pressler Street, Houston, Texas 77030, USA. ²⁶Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland 20892, USA. ²⁷Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA. ²⁸Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. ²⁹James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institutions, Baltimore, Maryland 21287, USA. ³⁰Cancer Prevention Institute of California, Fremont, California 94538; and Stanford University School of Medicine and Stanford Cancer Center, Stanford, California 94305, USA. ³¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, Maryland 21205, USA.

³²Department of Medicine, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ³³Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. ³⁴Department of Medicine, Division of Allergy, Pulmonary and Critical Care, 6100 Medical Center East, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8300, USA. ³⁵Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, and Center for Community Implementation and Dissemination Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ³⁶Department of Epidemiology, Gillings School of Global Public Health, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ³⁷Department of Urology, Northwestern University, Chicago, Illinois 60611, USA. ³⁸Department of Public Health Sciences, Henry Ford Hospital, Detroit, Michigan 48202, USA. ³⁹Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, 6701 Rockledge Drive, Bethesda, Maryland 20892, USA. ⁴⁰Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology & Health Services, University of Washington; Group Health Research Institute; Group Health Cooperative; 1730 Minor Ave., Seattle, Washington 98101, USA. ⁴¹Department of Epidemiology, University of Washington, Box 357236 Seattle, Washington 98195, USA. ⁴²Center for Public Health Genomics, University of Virginia, West Complex Room 6111, Charlottesville, Virginia 22908, USA. ⁴³Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. ⁴⁴Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. ⁴⁵Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland 20877, USA. ⁴⁶University of California San Francisco, San Francisco, California 94158, USA. ⁴⁷Institute for Human Genetics, Departments of Epidemiology and Biostatistics and Urology, University of California, San Francisco, San Francisco, California 94158, USA. ⁴⁸Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Wolstein Research Building, Cleveland, Ohio 44106, USA. ⁴⁹Brigham and Women's Hospital, Department of Medicine, Division of Sleep Medicine, 75 Francis Street, Boston, Massachusetts 02115, USA. ⁵⁰Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵¹Jackson State University, 1400 Lynch Street, Jackson, Mississippi 39217, USA. ⁵²Tougaloo College, 500 West County Line Road, Tougaloo, Mississippi 39174, USA. ⁵³Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁵⁴Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. ⁵⁵Department of Physiology and Biophysics, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA.

*These authors contributed equally to this work.

METHODS

Samples used for building the AA map. The 29,589 unrelated African-American samples derive from five sources. Informed consent was provided by all the individuals participating in the study, and was approved by all of the institutions responsible for sample collection.

The first source is the Candidate Gene Association Resource (CARE) study, a consortium of cohorts. We analysed CARE samples genotyped on the Affymetrix 6.0 array from the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). After removing individuals known to be related, and restricting to SNPs with good completeness in all cohorts, we had data from 6,209 individuals typed at 580,000 SNPs.

The second source consists of diverse studies carried out at the Children's Hospital of Philadelphia (CHOP), which has established a biobank for Philadelphia children to facilitate large genotype-phenotype association analysis. The cohort was recruited by CHOP clinicians, nursing and medical assistant staff within the CHOP Health Care Network, including primary care clinics and outpatient practices, from the hospital's patient base of over one million paediatric patients. All samples analysed here were genotyped on either the Illumina 610-Quad or Illumina HumanHap550 array. After removing individuals known to be related, identifying American Americans by multidimensional scaling on genotype data, and restricting to SNPs with a high level of completeness across samples, we had data from 7,503 samples typed at 491,572 SNPs.

The third source is the African American Breast Cancer Consortium (AABCC), consisting of the Multiethnic Cohort study (MEC), the Los Angeles component of the Women's Contraceptive and Reproductive Experiences study (CARE), the Women's Circle of Health Study (WCHS), the San Francisco Bay Area Breast Cancer study (SFBC), the Carolina Breast Cancer Study (CBCS), the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Cohort (PLCO), the Nashville Breast Health Study (NBHS) and the Wake Forest University Breast Cancer Study (WFBC), all genotyped on an Illumina 1M array. After data curation, including removal of samples with genetic evidence of being second-degree relatives or closer using the *smartrel* package of EIGENSOFT²⁶ (>0.2 correlation of genotype state), we had data from 5,203 women (about half cases and half controls) typed at 894,717 SNPs.

The fourth source is the African American Prostate Cancer Consortium (AAPCC), consisting of the MEC, the Southern Community Cohort Study (SCCS), PLCO, the Cancer Prevention Study II Nutrition Cohort (CPS-II), the Prostate Cancer Case-Control Studies at MD Anderson (MDA), the Identifying Prostate Cancer Genes study (IPCG), the Los Angeles Study of Aggressive Prostate Cancer (LAAPC), the Prostate Cancer Genetics Study (CaP Genes), the Case-Control Study of Prostate Cancer among African Americans in Washington DC (DCPC), the Gene-Environment Interaction in Prostate Cancer Study (GECAP) and the Cancer Prevention Study II (CPS-II), all typed on an Illumina 1M array. After the same data curation as the breast cancer study, we had data from 6,540 men (about half cases and half controls) typed at 896,036 SNPs.

The fifth source is individuals from the African American Lung Cancer Consortium (AALCC), including cases and controls from the MEC, the SCCS, PLCO, the MD Anderson (MDA) African American Lung Cancer Study, the NCI-Maryland Lung Cancer Case-Control Study, the University of California at San Francisco African American Lung Cancer Study and the Wayne State African American Lung Cancer Study, all genotyped on the Illumina 1M array. After data curation, we had data from 4,134 individuals typed at 906,687 SNPs.

Samples used for building the pedigree map. The pedigree map was built using data from 135 African-American nuclear families from CARE and 87 African-American families from CHOP for which genotyping data were available from at least two full siblings and at least one parent. The CARE studies that contributed samples were JHS (70 families, including 58 samples that we newly genotyped on the Affymetrix 6.0 array to increase the number of crossovers we could analyse) and CFS (65 families). For the families with a missing parent, we developed a Hidden Markov Model (HMM) approach to jointly estimate the genotype of the missing parent as well as to infer the position of crossover events in the offspring. The observed variables in the HMM were the genotypes of the available family members and the states of the HMM were the genotypes of the parents and the identity by descent (IBD) status of the children. A change in IBD status in an offspring is interpreted as a crossover event. Supplementary Note 2 provides details of the HMM used to infer positions of these pedigree crossover events.

Local ancestry inference and identification of crossover events. We merged the data for each cohort with phased YRI and CEU data from the HapMap3 data set²⁷. We filtered SNPs that had a frequency inconsistent with an 80–20% linear combination of YRI and CEU frequencies (t statistic with an absolute value of greater

than 3), potentially reflecting genotyping error in either the HapMap3 or the cohort data.

We ran HAPMIX on these data using a prior hypothesis of 20% European ancestry and 6 generations since mixture for each individual¹⁵. HAPMIX requires users to input a recombination map as a prior distribution, and we assumed that rates were constant across each chromosome arm with a total rate across each arm determined by the Rutgers genetic map⁶ (Supplementary Note 1).

Filtering of crossover events had three stages. First, we removed crossover events where the probability of occurrence was estimated to be less than 95% by HAPMIX. Second, we removed candidate crossover events that were non-monotonic, that is, where the probability of an overlapping crossover event with an ancestry switch in a different direction was $\geq 1\%$ within any inter-SNP interval. Third, we removed crossover events where either of the two flanking ancestry blocks was smaller than 2 cM in size as measured with respect to a published map based on linkage disequilibrium^{3,18} (Supplementary Note 1). For comparisons to the deCODE map and linkage-disequilibrium-based maps, we also removed segments of the genome within 5 Mb of the telomeres (to be consistent with the comparisons presented in the deCODE study where the same restriction was applied⁴).

Construction of the AA map. All 22 autosomes and chromosome X were split into approximately 1.3 million inter-SNP intervals based on the union of SNPs analysed across all five sample sets. Our goal was to estimate a crossover rate for each of these intervals. We modelled crossover rates such that the rate for each SNP interval is independent of every other SNP interval, motivated by a hotspot model. We used a gamma prior on rates with the mean estimated from the filtered HAPMIX output (Supplementary Note 1). We used a Gibbs sampler to sample rates in every SNP interval and to determine the location of a crossover event within the 95% range estimated by the HAPMIX output. In each round of the Gibbs sampler, we used the set of sampled rates in the previous round to construct a probability mass function for the SNP interval in which each crossover occurred, using an approach described in Supplementary Note 1 to approximate the probability mass function that HAPMIX would have produced conditional on the previous set of sampled rates. After sampling the location of the crossover events, we counted how many crossovers occurred in every SNP interval. We used these counts to construct a posterior distribution for the crossover rate in each SNP interval, taking advantage of the conjugacy of a Poisson likelihood and a gamma prior. We then sampled a crossover rate for each SNP interval from its respective gamma posterior distribution.

Candidate African-enriched hotspots. To identify candidate African-enriched hotspots, we used two pairs of maps: the previously available YRI map and CEU map, and the AE map and the S map. We combined information from both map pairs to enrich for regions with genuine differences between the West African and European populations. Specifically, we identified candidate hotspots as 2-kb intervals representing a peak in the AE map rate, where the estimated rate in the AE map was $>2 \text{ cM Mb}^{-1}$ and at least double that in the S map, and in addition the YRI map rate was $>2 \text{ cM Mb}^{-1}$ and at least double the CEU map rate. We took the resulting candidate hotspot set and defined hotspot boundaries by identifying the region flanking the 2 kb rate peak that had rates at least 50% of the peak value in the AE map. Regions larger than 5 kb were discarded. We similarly constructed a set of 'shared' hotspots but modified the initial criteria given the lack of obvious hotspots present only in people of European ancestry. Specifically, we identified 2 kb S map rate peak locations where both the S and CEU estimated rates were $>2 \text{ cM Mb}^{-1}$, while the AE and YRI map rates were below those in these respective European populations. We then narrowed the regions and filtered using the same procedure we had developed for the candidate African-enriched hotspots.

Association testing. MaCH²⁸ was used to impute up to 3,058,149 SNP genotypes from HapMap2 (ref. 18) into all African Americans we analysed, using the unrelated YRI and CEU samples as combined reference panels. We tested for association at all SNPs with minor allele frequency $> 1\%$. To restrict our analysis to individuals in whom the phenotype was measured accurately, we performed the association analysis with the AE and hotspot usage phenotypes only in individuals with at least 35 inferred crossovers. Association testing was carried out using linear regression, after controlling for gender, genome-wide European ancestry proportion (inferred by HAPMIX) and study (Supplementary Note 4). We observe slight inflation of the association statistics genome-wide compared with the expectation (the Genomic Control inflation factor²⁹ is 1.046 for the AE phenotype and 1.038 for the hotspot usage phenotype), which we propose may reflect cryptic relatedness among samples (Supplementary Note 4). We report P values after correction using Genomic Control²⁹.

Construction of PRDM9 tree. To examine the history of the PRDM9 ZF array and to place SNPs showing association with AE map usage within the framework of this history, we identified 19 SNPs from HapMap2 (ref. 18) that surrounded the

ZF array and that form a maximal block of SNPs where there is almost no evidence of recombination: $|D'| = 1$ for all pairs of SNPs in the data after removing 2 of 120 YRI and 1 of 120 CEU haplotypes (the chimpanzee genome was used to define the ancestral alleles). A unique 'gene tree' was then built, and we used *genetree*²⁵, which assumes a coalescent prior on genealogies, to approximately infer ages for these mutations conditional on the data (a caveat is that the tree building does not account for the HapMap SNP ascertainment scheme). Because *genetree* assumes a randomly mating population, and the YRI represent almost all the HapMap haplotype diversity in this region, we ran the software (2,000,000 importance samples, otherwise default parameters) on the YRI data only and used this to construct Fig. 2c. Each node of the tree corresponds to a unique haplotype at these 19 SNPs, whose frequency in both CEU and YRI is shown at the base of the figure.

Motif searching. We tested all candidate motifs of 5 to 9 base pairs for enrichment in our African-enriched hotspot set relative to our shared hotspot set. We counted occurrences of all tested motifs in repeat and non-repeat backgrounds separately, and computed a separate *P* value for each genomic background with a chi-squared test, based on a contingency table that compares the counts of a particular motif to the counts of all motifs of that size. We converted each *P* value to a *Z* score, added the scores on each background, and then obtained a corresponding combined *P* value. Motifs were considered statistically significant only if they passed four stringent criteria: (1) they were statistically significant after Bonferroni correction for the number of motifs tested; (2) they were overrepresented in the African-enriched set; (3) they were statistically significant on both the repeat and non-repeat backgrounds ($P < 0.01$) independently; and (4) they were statistically significant when the joint *P* value was calculated only by comparing the frequency of the motif to other motifs of identical G/C content (to eliminate false positives due to any difference in G/C content between the hotspot sets). This testing revealed a unique significant motif, the 9-nucleotide oligomer CCCCAGTGA. We explored whether flanking DNA around exact matches to this motif also had a role by testing whether bases at a given site relative to the motif were associated with the difference in rates between African- and European-ancestry populations (Kruskal–Wallis test). Rates were evaluated in the 2 kb surrounding each motif occurrence. We separately evaluated flanking sequence using both the difference between YRI/CEU map rates, and the difference between the AE/S map rates, leading to the identification of the 17-bp

consensus African-enriched motif (Supplementary Note 6 has full details). To identify close matches to this 17-bp motif among all matches to the 9-bp motif in the genome, for every occurrence of the 9-bp motif, we scored the flanking sequence bases proportionately to the relative increase in average crossover rate difference associated with each base, then multiplied across bases in the 17-mer region to provide an overall score. We ranked occurrences according to this score, and plotted rates around the top 500 (Fig. 3b). We verified these findings by measuring average crossover differences for each base using only odd chromosomes and used these to score motif occurrences on the (non-overlapping) set of even chromosomes, and vice versa (Supplementary Fig. 8).

PRDM9 ZF length typing and genotyping of rs6889665. To determine the number of ZF motifs of *PRDM9* in a subset of the samples used to build the map, published primer pairs⁴ were used to amplify this region (forward: 5'-GGCCAGAAAGTGAATCCAGG-3', reverse: 5'-GGGGAATATAAGGGGTCAGC-3'). Product lengths ranged between 7 and 20 repeats (801–1,893 bp). Four of the 166 African-American samples did not show an amplification product, presumably because of insufficient DNA quality. We also genotyped 90 YRI and 90 CEU HapMap samples.

The SNP rs6889665 was genotyped in the same samples using an allelic discrimination assay (forward primer: 5'-aaacttggaacatccatagggt-3', reverse primer: 5'-cgaaaggagaaagcataatcc-3', Locked Nucleic Acid (LNA) probe 'C': 5'-/6-FAM/aGGGatAaatgaag/BHQ/-3', LNA-probe 'T': 5'-/HEX/AGAGatAaatGaagg/BHQ/-3'; LNA bases are given in capital letters). Reporter dyes: 6-FAM, 6-carboxyfluorescein; HEX, hexachlorofluorescein. Quencher: BHQ, Black Hole Quencher 1. Only one out of the 166 African-American samples failed in this assay. The same YRI and CEU samples as above were also genotyped.

26. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
27. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
28. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

Hydrogen is an energy source for hydrothermal vent symbioses

Jillian M. Petersen^{1*}, Frank U. Zielinski^{1,2*}, Thomas Pape³, Richard Seifert⁴, Cristina Moraru¹, Rudolf Amann¹, Stephane Hourdez⁵, Peter R. Girguis⁶, Scott D. Wankel⁶, Valerie Barbe⁷, Eric Pelletier⁷, Dennis Fink¹, Christian Borowski¹, Wolfgang Bach⁸ & Nicole Dubilier¹

The discovery of deep-sea hydrothermal vents in 1977 revolutionized our understanding of the energy sources that fuel primary productivity on Earth. Hydrothermal vent ecosystems are dominated by animals that live in symbiosis with chemosynthetic bacteria. So far, only two energy sources have been shown to power chemosynthetic symbioses: reduced sulphur compounds and methane. Using metagenome sequencing, single-gene fluorescence *in situ* hybridization, immunohistochemistry, shipboard incubations and *in situ* mass spectrometry, we show here that the symbionts of the hydrothermal vent mussel *Bathymodiolus* from the Mid-Atlantic Ridge use hydrogen to power primary production. In addition, we show that the symbionts of *Bathymodiolus* mussels from Pacific vents have *hupL*, the key gene for hydrogen oxidation. Furthermore, the symbionts of other vent animals such as the tubeworm *Riftia pachyptila* and the shrimp *Rimicaris exoculata* also have *hupL*. We propose that the ability to use hydrogen as an energy source is widespread in hydrothermal vent symbioses, particularly at sites where hydrogen is abundant.

Deep-sea hydrothermal vents and their associated chemosynthetic communities were discovered in 1977¹. Since then, two energy sources have been shown to fuel primary production by the symbiotic bacteria that form the basis of the food chain in marine chemosynthetic ecosystems. In 1981, chemolithoautotrophic bacteria that use reduced sulphur compounds as an energy source were discovered in the gutless *Riftia pachyptila* tubeworms from hydrothermal vents in the Pacific^{2,3}. Five years later, the first symbionts that use methane as an energy source were discovered in mussels from hydrocarbon seeps in the Gulf of Mexico^{4,5}. Since then, a vast array of chemosynthetic ecosystems has been explored, and novel symbioses of phylogenetically diverse hosts and symbionts are constantly being described⁶. Despite this, no other source of energy for metazoan chemosynthetic symbioses has yet been found. This is remarkable, given that many potential sources of energy for chemosynthesis are available, such as hydrogen, ammonium, ferrous iron and manganese(II)⁷, and free-living vent microbes able to use these energy sources are well known^{8–10}.

Some hydrothermal vents produce fluids with very high hydrogen concentrations due to the interaction of seawater with mantle-derived ultramafic rocks. Fluids originating from ultramafic-hosted vents are also characterized by high methane concentrations, whereas H₂S concentrations are rather low. In contrast, basalt-hosted vents produce fluids comparably high in H₂S but low in H₂ and CH₄. The Logatchev vent field at 14° 45' N on the Mid-Atlantic Ridge (MAR) is located in a ridge segment characterized by ultramafic outcrops (Supplementary Fig. 2). Fluids venting at this site have the highest hydrogen concentrations ever measured in hydrothermal systems (19 mM in the end-member fluids)¹¹, which would provide a rich source of energy for chemosynthetic microbes. Hydrogen is a particularly favourable electron donor as the energy yield from hydrogen oxidation is much higher than from methane oxidation, sulphur oxidation and all other potential electron donors for chemolithoautotrophic growth under

standard conditions¹². In fact, our thermodynamic model predicts that at the Logatchev vent field, aerobic hydrogen oxidation could provide up to 7 times more energy per kilogram of vent fluid than methane oxidation, and up to 18 times more energy per kilogram of vent fluid than sulphide oxidation (Supplementary Fig. 1).

At Logatchev, *Bathymodiolus puteoserpentis* mussels are by far the most abundant macrofauna¹³. They live in a dual symbiosis with methane-oxidizing and chemoautotrophic sulphur-oxidizing bacteria that are hosted in their gills^{14–16}. Here we show that the sulphur-oxidizing symbiont of *B. puteoserpentis* uses hydrogen as an energy source. We also show that *Bathymodiolus* symbionts from basalt-hosted MAR vent fields can oxidize hydrogen. This ability is therefore not limited to the ultramafic-hosted vent fields and could also be powering chemosynthetic symbioses at basalt-hosted vent fields.

Uptake hydrogenase genes in mussels

The key enzymes involved in hydrogen metabolism are hydrogenases, which catalyse the reaction: $\text{H}_2 \leftrightarrow 2\text{H}^+ + 2\text{e}^-$ (ref. 17). Enzymes of the group 1 NiFe hydrogenases are membrane-bound respiratory enzymes that channel electrons from hydrogen into the quinone pool, providing the link between hydrogen oxidation and energy production^{17,18}. The large subunit of the membrane-bound uptake hydrogenase is encoded by the *hupL* gene. We amplified and sequenced this gene from symbiont-containing *B. puteoserpentis* gill tissues from Logatchev (Supplementary Fig. 3). The closest related sequence was the large subunit of the NiFe hydrogenase from the alphaproteobacterium *Oligotropha carboxidovorans* (79.6% amino acid identity). *O. carboxidovorans* can grow chemolithoautotrophically under aerobic conditions using either CO or H₂ as an electron donor¹⁹. Phylogenetic analyses based on multiple treeing methods placed the enzyme from *B. puteoserpentis* in a well-supported cluster with other group 1 NiFe hydrogenases, showing the genetic potential for hydrogen oxidation

¹Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany. ²Helmholtz Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, Germany. ³MARUM - Center for Marine Environmental Sciences and Department of Geosciences, University of Bremen, Klagenfurter Strasse, 28334 Bremen, Germany. ⁴University of Hamburg, Institute for Biogeochemistry and Marine Chemistry, Bundesstrasse 55, 20146 Hamburg, Germany. ⁵Equipe Genetique des Adaptations aux Milieux Extrêmes, CNRS-UPMC UMR 7144, Station Biologique, BP74, 29682 Roscoff, France.

⁶Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁷Commissariat à l'Energie Atomique/Genoscope, 91000 Évry, France, Centre National de la Recherche Scientifique, UMR8030, 91000 Évry, France, and Université d'Évry Val d'Essonne 91000 Évry, France. ⁸Department of Geosciences, University of Bremen, PO Box 33 04 40, 28334 Bremen, Germany.

*These authors contributed equally to this work.

linked to energy generation in *B. puteoserpentis* endosymbionts. To examine whether the endosymbionts of mussels from hydrogen-poor habitats also have the genetic potential for hydrogen uptake, we tried to amplify the *hupL* gene from gill tissues of mussels from basalt-hosted vents and cold seeps that have fluids with low hydrogen concentrations. Indeed, we could amplify the *hupL* gene from mussels from basalt-hosted vents, including undescribed *Bathymodiulus* mussels (*Bathymodiulus* spp.) from vents on the southern MAR (Wideawake at 4° 48' S, 5,600 km from Logatchev, and Lilliput at 9° 33' S, 6,500 km from Logatchev), and *B. aff. thermophilus* from the Axial Dome vent on the Pacific–Antarctic Ridge at 37° 47' S (Supplementary Figs 2, 3 and Supplementary Table 7), showing that the genetic potential for hydrogen oxidation is not restricted to mussel symbionts from ultramafic-hosted vent fields that have high hydrogen concentrations. Intriguingly, we could not amplify this gene from any of the cold seep mussels investigated (see Supplementary Discussion).

B. puteoserpentis symbionts use H₂

Given the genetic potential for hydrogen uptake in *Bathymodiulus* mussels, we incubated *B. puteoserpentis* mussel gill tissues from the ultramafic-hosted Logatchev vent field on the MAR with hydrogen at partial pressures of ~100 p.p.m. in the headspace (0.08 μM H₂ dissolved in the medium) and measured its consumption over time. Hydrogen was taken up rapidly by the symbiont-containing gill tissues at a rate of $\sim 650 \pm 200 \text{ nmol h}^{-1} (\text{g wet weight})^{-1}$ ($n = 7$). In contrast, symbiont-free foot tissue did not consume hydrogen at rates above the negative controls—boiled gill tissue and seawater (Fig. 1 and Supplementary Table 3).

Because hydrogen uptake by bacteria is not necessarily coupled to CO₂ fixation^{17,20}, we incubated mussel gill tissues with hydrogen in seawater containing ¹⁴C-bicarbonate to determine whether hydrogen is an energy source for autotrophic CO₂ fixation by *B. puteoserpentis* symbionts. Control gill tissues were incubated in the presence of sulphide or without an electron donor. ¹⁴C uptake was stimulated by sulphide, known to be an energy source for the sulphur-oxidizing symbionts of *Bathymodiulus* mussels²¹, and also by hydrogen (Fig. 1). The rates of carbon fixation with hydrogen were comparable to those supported by sulphide oxidation, which suggests that hydrogen could be fuelling autotrophy to the same extent as sulphide. We therefore conclude that hydrogen provides energy for the production of mussel biomass at the Logatchev hydrothermal vent field.

Symbionts from basalt-hosted vents use H₂

We examined whether the symbionts of *Bathymodiulus* mussels from basalt-hosted vent fields on the southern MAR at Comfortless Cove and Lilliput with low *in situ* hydrogen concentrations could also consume hydrogen. Hydrogen concentrations measured in discrete samples of the diffuse fluids from these vents were typically below 0.1 μM (refs 22,23) as opposed to Logatchev, which had up to 154 μM , on the basis of discrete sampling (Supplementary Table 3). We incubated gill tissues of *B. spp.* mussels and measured hydrogen consumption over time as described earlier. Our results show that mussels from these vents were also able to take up hydrogen (Fig. 1 and Supplementary Table 3). However, the rates at which the symbiotic gill tissues of the southern MAR mussels consumed hydrogen were 20- to 30-fold lower than those of mussels from Logatchev.

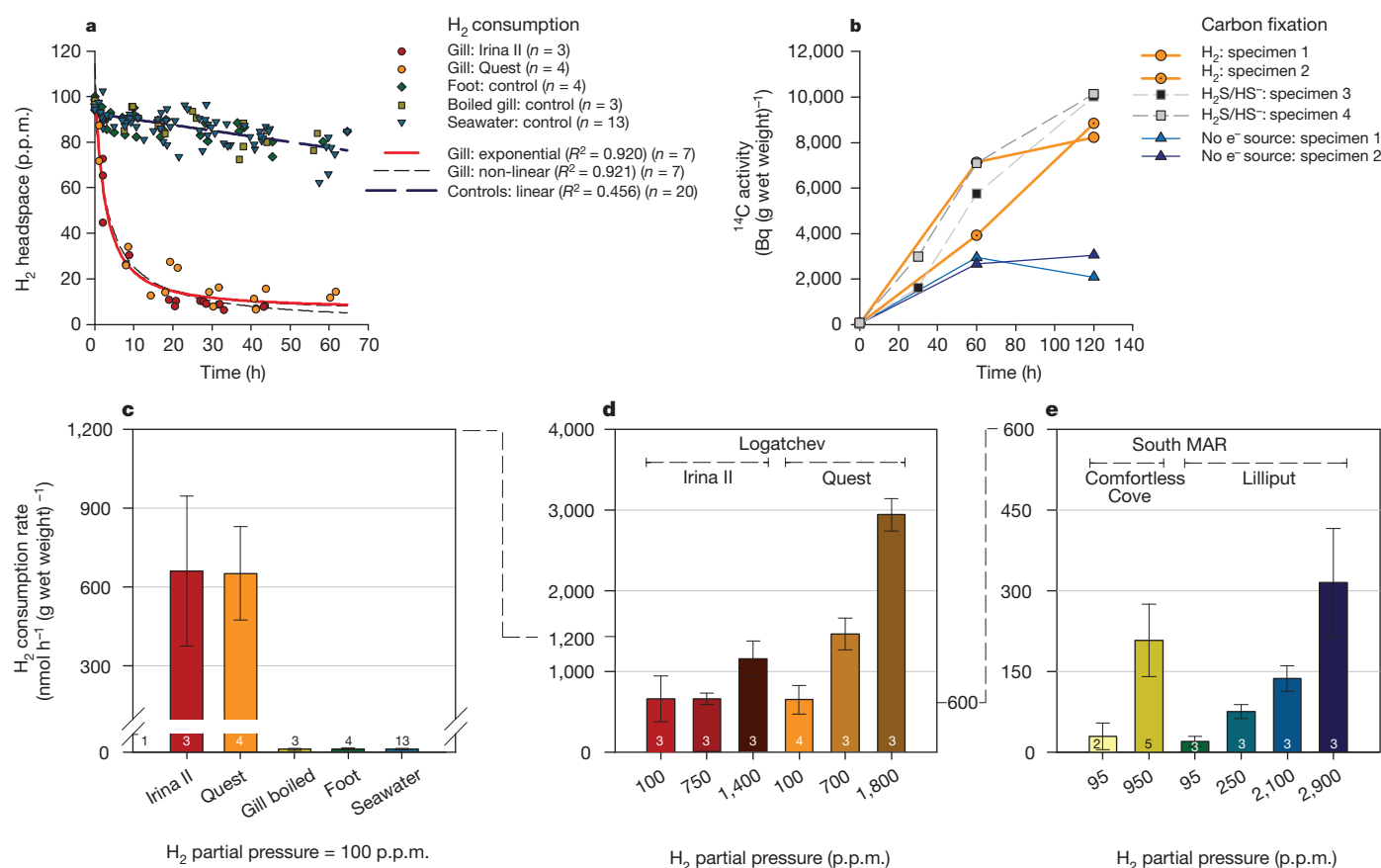


Figure 1 | Hydrogen consumption in *Bathymodiulus* gills. **a**, **b**, Hydrogen was consumed rapidly by *B. puteoserpentis* gills (Logatchev) and carbon fixation was stimulated after incubation with ¹⁴C-bicarbonate in the presence of hydrogen or sulphide. Each data point is one distinct measurement. **c**, Consumption rates of gills from two sites at Logatchev were in the same

range. **d**, **e**, Rates from Logatchev and the southern MAR fields (*B. spp.*; Comfortless Cove, 4° 48' S and Lilliput, 9° 33' S) increased with increasing hydrogen partial pressures. Rates were roughly 20–30-fold higher at the hydrogen-rich Logatchev field than at the hydrogen-poor southern MAR vents. Error bars represent one standard deviation.

Effect of H₂ concentration on consumption

Some hydrogen-oxidizing microorganisms are known to induce hydrogenase expression only in the presence of hydrogen, and some express this enzyme constitutively at a low level, but increase its expression upon incubation with hydrogen²⁴. To investigate the effect of dissolved hydrogen concentration on hydrogen consumption rates, we incubated gill tissues from the Logatchev and southern MAR venting sites at partial pressures of up to 3,000 p.p.m. ($\sim 2.3 \mu\text{M}$ dissolved H₂). Hydrogen consumption rates in mussels from Logatchev, Comfortless Cove and Lilliput increased by 135, 21 and 8 nmol h⁻¹ (g wet weight)⁻¹, respectively, for each 100 p.p.m. increase in partial pressure (Fig. 1). Thus, hydrogen uptake is clearly stimulated by increasing hydrogen concentrations.

The sulphur-oxidizing symbiont uses H₂

Many phylogenetically diverse microorganisms can grow autotrophically on hydrogen. *Bathymodiolus* mussels from the MAR host two types of chemosynthetic gammaproteobacteria: a sulphur oxidizer and a methane oxidizer¹⁴ (Fig. 2). In addition, these mussels have a gammaproteobacterial parasite that infects gill nuclei²⁵. Hydrogen is known to provide energy for chemo- and photolithoautotrophic

growth of some free-living sulphur-oxidizing bacteria^{26,27}. Some free-living methane-oxidizing bacteria can also oxidize hydrogen in addition to methane²⁸. Lastly, hydrogen has been shown to be a key determinant of the virulence of pathogenic *Helicobacter pylori*²⁹. Accordingly, it was unclear which of the three types of bacteria associated with *Bathymodiolus* mussels could be using the hydrogen. To investigate this further, we used three molecular methods allowing us to link identity with function at the DNA and protein level: genome sequencing of the mussel symbionts, single-gene fluorescence *in situ* hybridization (geneFISH) and immunohistochemistry, the latter two combined with 16S rRNA FISH.

In the genomes of hydrogen-oxidizing bacteria, the genes for the large and small subunits of membrane-bound NiFe hydrogenases are often found clustered together with the genes necessary for biosynthesis, maturation and processing¹⁷. We used metagenomics to investigate hydrogen oxidation in the genome of *Bathymodiolus* symbionts from the southern MAR (Lilliput). We found a hydrogenase operon with 19 open reading frames on a 726-kb genome fragment (Supplementary Fig. 4). The large subunit gene from this genome fragment was 99.6% identical to the *hupL* gene amplified from the Wideawake (southern MAR) mussels using polymerase chain reaction (PCR), and 83% identical to the gene from *B. puteoserpentis* from Logatchev on the northern MAR. Homologues of all structural genes and those necessary for hydrogenase synthesis, assembly and function in the chemolithoautotrophic hydrogen oxidizer *Cupriavidus necator*³⁰ were present on our genome fragment. The southern MAR *Bathymodiolus* symbiont therefore has all of the genetic components required for hydrogen uptake. On the same genome fragment, we found the key genes for sulphur oxidation via the reverse dissimilatory sulphite reductase (rDsr) and Sox pathways, as well as for CO₂ fixation by the Calvin–Benson–Bassham (CBB) cycle (Supplementary Fig. 4), indicating that this fragment is from the sulphur-oxidizing symbiont of *Bathymodiolus*.

To establish further that it is the sulphur-oxidizing symbiont that uses hydrogen, we used geneFISH, a method in which simultaneous detection of single genes and rRNA enables the linking of function and identity at the single-cell level³¹. We detected the *hupL* gene in the symbiont-containing gill tissue of *B. puteoserpentis* (Fig. 2). GeneFISH signals from probes designed to target the *hupL* gene amplified by PCR from *B. puteoserpentis* overlapped with the 16S rRNA signals from the previously described sulphur-oxidizing symbiont (Fig. 2). This further suggests that the *hupL* gene we amplified is from the sulphur-oxidizing symbiont of *B. puteoserpentis*.

To show that the sulphur-oxidizing symbiont expresses the uptake hydrogenase, we combined immunohistochemistry with FISH of the 16S rRNA of the symbionts of *B. puteoserpentis*, thus linking identity and function at the level of gene expression. Using a polyclonal anti-serum against the membrane-bound hydrogen-uptake NiFe hydrogenase from *C. necator* (74% amino acid sequence identity to the *B. puteoserpentis* symbiont HupL; Supplementary Fig. 5), we were able to detect hydrogenase expression in single symbiont cells in *B. puteoserpentis* (Fig. 2). Signals from the anti-hydrogenase antibody were seen in gill bacteriocytes of *B. puteoserpentis* where they overlapped with the FISH signals from the sulphur-oxidizing symbiont, but not with those from the methane-oxidizing symbiont.

Environmental significance of H₂ use

One of the major challenges in hydrothermal vent research is to link physiological experiments, performed on board the research vessel at atmospheric pressure, with processes occurring under *in situ* temperature and hydrostatic pressure. We deployed an *in situ* mass spectrometer (ISMS) to measure hydrogen concentrations in the mussel habitat at Logatchev, and compared these with temperature as a conservative tracer for mixing of the hydrothermal fluids with seawater. Simultaneous measurements of hydrogen concentrations and temperature were made in two different settings; first, in an area of

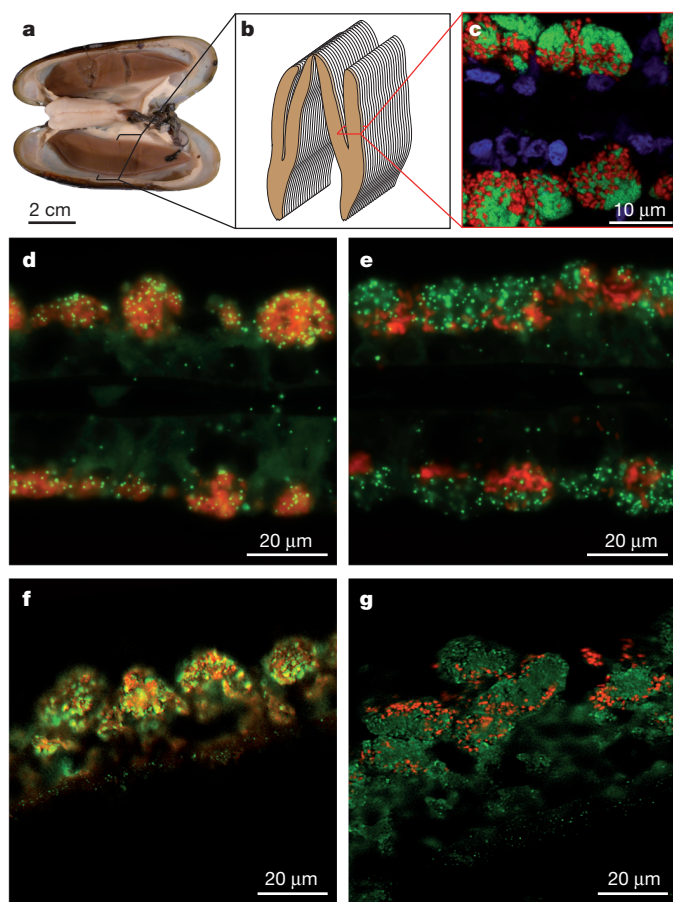


Figure 2 | The sulphur-oxidizing symbiont has the gene for hydrogen uptake, which it expresses. **a**, *B. puteoserpentis* with brown, symbiont-containing gills. **b**, Morphology of a single gill, each composed of many filaments (modified from ref. 48). **c**, FISH of a cross-section through a single gill filament (host cell nuclei, blue; sulphur-oxidizing symbiont, green; methane-oxidizing symbiont, red). **d**, **e**, *hupL* geneFISH (signals in green) with 16S rRNA FISH (signals in red) for the sulphur-oxidizing symbiont (**d**) and the methane-oxidizing symbiont (**e**). The negative control gave no gene signals (Supplementary Fig. 7). **f**, **g**, HupL immunohistochemistry (signals in green) with 16S rRNA FISH (signals in red) for the sulphur-oxidizing symbiont (**f**) and the methane-oxidizing symbiont (**g**). Yellow colours in **d** and **f** appear where green and red signals overlap.

focused flow, directly at the source where the fluids exit the seafloor. Here, the fluids had not been exposed to *B. puteoserpentis* mussels or other macrofauna. Second, we measured hydrogen concentrations and temperature in an area where the fluids had passed through a *B. puteoserpentis* mussel bed. The slope of the regression line calculated for hydrogen concentration versus temperature was significantly lower in the mussel bed compared to the source fluid ($P < 0.0001$), indicating that fluids in the mussel bed are hydrogen-depleted compared to the source fluid (Fig. 3). This difference is probably due to hydrogen consumption by the sulphur-oxidizing symbionts of *B. puteoserpentis*. We observed the same effect with measurements of methane and temperature ($P < 0.0001$; Supplementary Fig. 6). These measurements confirm that significant amounts of hydrogen are being consumed in Logatchev mussel beds.

The terrestrial hydrogen biogeochemical cycle has been the topic of numerous studies (summarized in ref. 32), and interest in the global hydrogen cycle has recently grown with the prospect of a hydrogen economy³³. In contrast to the terrestrial environment, there is a paucity of data on hydrogen turnover in the oceans. Conservative calculations based on our measurements of uptake rates in Logatchev mussels show that at 50 μM dissolved hydrogen a single mussel with a gill weight of 5 g could oxidize up to 435 $\mu\text{mol H}_2 \text{ h}^{-1}$ (Supplementary Information). On the basis of our own estimates and those of others, the mussel population at Logatchev spreads in layers over at least 200 m^2 and accounts for most of the invertebrate biomass¹³. One square meter is covered by approximately 1,100–2,500 adult individuals corresponding to a total population of 250,000 to 500,000 individuals. This population could consume between 0.5 and 1.0 $\text{mol H}_2 \text{ m}^{-2} \text{ h}^{-1}$, that is, in total up to 200 $\text{mol H}_2 \text{ h}^{-1}$. Previous studies have reported hydrogen uptake rates in the range of several $\text{nmol ml}^{-1} \text{ h}^{-1}$ in hydrothermal fluids³⁴, and several $\text{nmol m}^{-3} \text{ h}^{-1}$ in coastal waters³⁵. On the basis of the limited data available for comparison, the symbionts of *Bathymodiolus* mussels from the MAR probably have a role as a significant hydrogen sink at these hydrothermal vents.

We do not know if the chemoautotrophic symbionts of vent mussels from other mid-ocean ridges are actively consuming hydrogen, although we show here that *B. aff. thermophilus* from the Pacific–Antarctic Ridge has the key gene necessary. Hydrogen is clearly present in the fluids at many basalt-hosted vents^{36,37}, and could well be powering biomass production at these sites as well. In addition, use of hydrogen as an energy source may also have a role in other chemoautotrophic

symbioses; the episymbionts of the hydrothermal vent shrimp *Rimicaris exoculata* from MAR vent fields have *hupL* genes³⁸ (Supplementary Fig. 3), as does the endosymbiont of the giant tubeworm *Riftia pachyptila* from basalt-hosted vents in the East Pacific, which it expresses *in situ* (S. Markert, personal communication). This indicates that hydrogen use could be widespread in chemosynthetic symbioses.

METHODS SUMMARY

A detailed description of the sampling sites (Supplementary Tables 1 and 2) and all methods used in this study can be found in the Supplementary Information.

For amplification of the *hupL* gene, we extracted DNA from gill samples³⁹ and used published primers⁴⁰. Cloning and sequencing were done as described elsewhere⁴¹. Maximum likelihood and parsimony *HupL* phylogenies were calculated in ARB⁴² using a MAFFT⁴³ alignment. For incubation experiments, *Bathymodiolus* mussels were collected from hydrothermal vent fields on the MAR by remotely operated vehicles and incubated immediately after recovery on board the research vessel. Mussel gill and foot tissues were incubated in sterile-filtered seawater in glass serum vials (Supplementary Table 4). Hydrogen was added to the vials, and the change in headspace concentration was measured over time by gas chromatography. CO_2 fixation was measured by incubation of gill tissues with ^{14}C -bicarbonate after adding H_2 , Na_2S (for $\text{H}_2\text{S}/\text{HS}^-$), or no electron donor. The incorporation of labelled ^{14}C uptake over time was measured by liquid scintillation counting in separate tissue pieces from the same individual that were incubated for 0, 30, 60 and 120 min. Mussels for genome sequencing were collected from the Lilliput vent site on the MAR. Gill tissue was immediately homogenized and fixed on board. We enriched for symbiont cells by filtering sequentially through 8, 5, 3 and 2 μm GTTP filters. We extracted DNA³⁹ and constructed a 6 kb paired-end library, which was sequenced by 454-Titanium and Illumina. Titanium reads were assembled using Newbler (v.2.3-PreRelease-10/19/2009-454; Life Sciences Corporation). One 726 kb scaffold was identified that contained the hydrogenase operon, and this scaffold was annotated by RAST⁴⁴. Polynucleotide probe design and geneFISH were done as previously described^{31,45}, with modifications listed in Supplementary Information (Supplementary Tables 5 and 6). Immunohistochemistry was done by microwaving gill sections of *B. puteoserpentis*, blocking in western blocking reagent (Roche), incubating in primary antibody, then in HRP-conjugated secondary antibody. Signal amplification was done by CARD⁴⁶. The ISMS was deployed as previously described⁴⁷ with modifications listed in Supplementary Information. Statistical analysis was done with the software JMP5.

Received 15 April; accepted 20 June 2011.

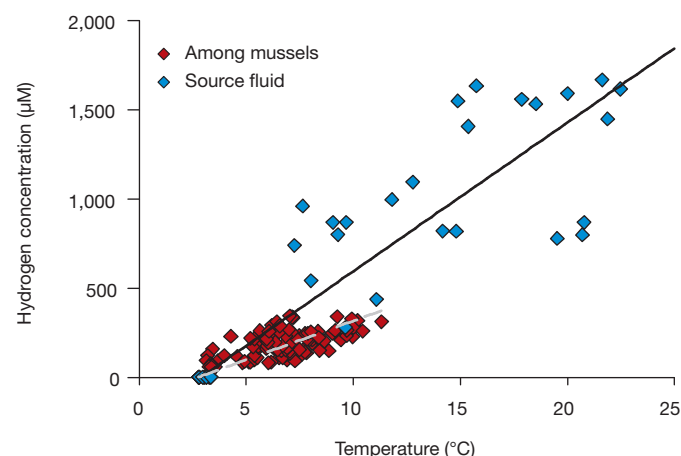


Figure 3 | Hydrogen is consumed in mussel beds of *B. puteoserpentis*. Plot of hydrogen concentrations versus temperature, measured by *in situ* mass spectrometry in the source fluid and in a *B. puteoserpentis* mussel bed. The unbroken line shows linear regression analysis for the source fluid ($y = 83.308x - 240$, $R^2 = 0.63$), the broken line shows linear regression analysis for the mussel bed fluid ($y = 43.576x - 120$, $R^2 = 0.56$). The difference in the slopes of the regression lines between the source fluid and in the mussel bed is due to the consumption of hydrogen in the mussel bed.

- Corliss, J. B. *et al.* Submarine thermal springs in the Galapagos Rift. *Science* **203**, 1073–1083 (1979).
- Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science* **213**, 340–342 (1981).
- Felbeck, H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213**, 336–338 (1981).
- Childress, J. J. *et al.* A methanotrophic marine molluscan (Bivalvia, Mytilidae) symbiosis: mussels fueled by gas. *Science* **233**, 1306–1308 (1986).
- Cavanaugh, C. M., Levering, P. R., Maki, J. S., Mitchell, R. & Lidstrom, M. E. Symbiosis of methylotrophic bacteria and deep-sea mussels. *Nature* **325**, 346–348 (1987).
- Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Rev. Microbiol.* **6**, 725–740 (2008).
- Tivey, M. K. Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography (Wash. D.C.)* **20**, 50–65 (2007).
- Fisher, C. R., Takai, K. & Le Bris, N. Hydrothermal vent ecosystems. *Oceanography (Wash. D.C.)* **20**, 14–23 (2007).
- Takai, K., Nakagawa, S., Reysenbach, A.-L. & Hock, J. In *Back-Arc Spreading Systems—Geological, Biological, Chemical, and Physical Interactions* (eds Christie, D. M. *et al.*) 185–213 (American Geophysical Union, 2006).
- Perner, M. *et al.* The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15° N on the Mid-Atlantic Ridge. *FEMS Microbiol. Ecol.* **61**, 97–109 (2007).
- Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L. M. & Seifert, R. Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15° N on the Mid-Atlantic Ridge: temporal and spatial investigation. *Chem. Geol.* **242**, 1–21 (2007).
- Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25**, 175–243 (2001).
- Gebruk, A. V., Chevaldonné, P., Shank, T., Lutz, R. A. & Vrijenhoek, R. C. Deep-sea hydrothermal vent communities of the Logatchev area (14° 45' N, Mid-Atlantic Ridge): diverse biotopes and high biomass. *J. Mar. Biol. Assoc. UK* **80**, 383–393 (2000).
- Duperron, S. *et al.* A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from

- hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ. Microbiol.* **8**, 1441–1447 (2006).
15. Petersen, J. M. & Dubilier, N. Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* **1**, 319–335 (2009).
 16. Wendeberg, A., Zielinski, F. U., Borowski, C. & Dubilier, N. Expression patterns of mRNAs for methanotrophy and thiotrophy in symbionts of the hydrothermal vent mussel *Bathymodiolus puteoserpentis*. *ISME J.* doi:10.1038/ismej.2011.81 (7 July 2011).
 17. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
 18. Bernhard, M., Schwartz, E., Rietdorf, J. & Friedrich, B. The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling. *J. Bacteriol.* **178**, 4522–4529 (1996).
 19. Meyer, O. & Schlegel, H. G. Reisolation of the carbon monoxide utilizing hydrogen bacterium *Pseudomonas carboxydovorans* (Kistner) comb. nov. *Arch. Microbiol.* **118**, 35–43 (1978).
 20. Schwartz, E. & Friedrich, B. in *The Prokaryotes: A Handbook on the Biology of Bacteria* (eds Dworkin, M. et al.) Vol. 2, 496–563 (Springer, 2006).
 21. Nelson, D. C., Hagan, K. D. & Edwards, D. B. The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar. Biol.* **121**, 487–495 (1995).
 22. Haase, K. M. et al. Diking, young volcanism and diffuse hydrothermal activity on the southern Mid-Atlantic Ridge: the Lilliput field at 9° 33' S. *Mar. Geol.* **266**, 52–64 (2009).
 23. Haase, K. M. et al. Young volcanism and related hydrothermal activity at 5° S on the slow-spreading southern Mid-Atlantic Ridge. *Geochem. Geophys. Geosys.* **8**, Q11002 (2007).
 24. Friedrich, B. & Schwartz, E. Molecular biology of hydrogen utilization in aerobic chemolithotrophs. *Annu. Rev. Microbiol.* **47**, 351–383 (1993).
 25. Zielinski, F. U. et al. Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ. Microbiol.* **11**, 1150–1167 (2009).
 26. Ohmura, N., Sasaki, K., Matsumoto, N. & Saiki, H. Anaerobic respiration using Fe³⁺, S⁰, and H₂ in the chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans*. *J. Bacteriol.* **184**, 2081–2087 (2002).
 27. Imhoff, J. F., Hiraishi, A. & Söhling, J. in *Bergey's Manual of Systematic Bacteriology* (eds Brenner, D. J. et al.) Vol. 2, part A, 119–132 (Springer, 2005).
 28. DiSpirito, A. A., Kunz, R. C., Choi, D.-W. & Zahn, J. A. in *Respiration in Archaea and Bacteria* (ed. Zannoni, D.) Vol. 2, 149–168 (Springer, 2004).
 29. Olson, J. W. & Maier, R. J. Molecular hydrogen as an energy source for *Helicobacter pylori*. *Science* **298**, 1788–1790 (2002).
 30. Bowien, B. & Schlegel, H. G. Physiology and biochemistry of aerobic hydrogen-oxidizing bacteria. *Annu. Rev. Microbiol.* **35**, 405–452 (1981).
 31. Moraru, C., Lam, P., Fuchs, B. M., Kuypers, M. M. M. & Amann, R. GeneFISH—an *in situ* technique for linking gene presence and cell identity in environmental microorganisms. *Environ. Microbiol.* **12**, 3057–3073 (2010).
 32. Constant, P., Piossant, L. & Villemur, R. Tropospheric H₂ budget and the response of its soil uptake under the changing environment. *Sci. Total Environ.* **407**, 1809–1823 (2009).
 33. Tromp, T., Shia, R.-L., Allen, M., Eiler, J. M. & Yung, Y. L. Potential environmental impact of a hydrogen economy on the stratosphere. *Science* **300**, 1740–1742 (2003).
 34. Perner, M., Petersen, J. M., Zielinski, F., Gennerich, H. H. & Seifert, R. Geochemical constraints on the diversity and activity of H₂-oxidizing microorganisms in diffuse hydrothermal fluids from a basalt- and an ultramafic-hosted vent. *FEMS Microbiol. Ecol.* **74**, 55–71 (2010).
 35. Punshon, S., Moore, R. M. & Xie, H. Net loss rates and distribution of molecular hydrogen (H₂) in mid-latitude coastal waters. *Mar. Chem.* **105**, 129–139 (2007).
 36. Welhan, J. A. & Craig, H. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A. et al.) 391–410 (Plenum, 1983).
 37. Lilley, M. D., DeAngelis, M. A. & Gordon, L. I. CH₄, H₂, CO and N₂O in submarine hydrothermal vent waters. *Nature* **300**, 48–50 (1982).
 38. Hügler, M., Petersen, J. M., Dubilier, N., Imhoff, J. F. & Sievert, S. M. Pathways of carbon and energy metabolism of the epibiotic community associated with the deep-sea hydrothermal vent shrimp *Rimicaris exoculata*. *PLoS ONE* **6**, e16018 (2011).
 39. Zhou, J. Z., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
 40. Csáki, R., Hanczár, T., Bodrossy, L., Murrell, J. C. & Kovács, K. L. Molecular characterization of structural genes coding for a membrane bound hydrogenase in *Methylococcus capsulatus* (Bath). *FEMS Microbiol. Lett.* **205**, 203–207 (2001).
 41. Petersen, J. M. et al. Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12**, 2204–2218 (2010).
 42. Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
 43. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 39–64 (2009).
 44. Aziz, R. K. et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
 45. Moraru, C., Moraru, G., Fuchs, B. M. & Amann, R. Concepts and software for a rational design of polynucleotide probes. *Environ. Microbiol. Rep.* **3**, 69–78 (2011).
 46. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68**, 3094–3101 (2002).
 47. Wankel, S. D. et al. Influence of subsurface biosphere on geochemical fluxes from diffuse hydrothermal fluids. *Nature Geosci.* **4**, 461–468 (2011).
 48. Le Pennec, M. & Hily, A. Anatomie, structure et ultrastructure de la branchie d'un Mytilidae des sites hydrothermaux du Pacifique oriental. *Oceanol. Acta* **7**, 517–523 (1984).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the chief scientists, and the captains and crews of the research vessels and remotely operated vehicles involved in sampling and analyses at sea. Thank you to D. Garbe-Schönberg, K. van der Heijden and J. Stecher for on-board sampling and analysis, and S. Duperron, M.-A. Cambon-Bonavita and M. Zbinden for providing samples. We acknowledge B. Friedrich and O. Lenz for the antiserum against the *C. necator* uptake hydrogenase, J. Milucka for help with western blots and T. Holler for culturing *C. necator*. S. Wetzel provided technical assistance. This work was supported by the German Science Foundation (DFG) Priority Program 1144 “From Mantle to Ocean: Energy-, Material- and Life Cycles at Spreading Axes” (publication number 60), the DFG Cluster of Excellence “The Ocean in the Earth System” at MARUM (Center for Marine Environmental Sciences), and the Max Planck Society.

Author Contributions J.M.P., F.U.Z., T.P., R.S., C.B., D.F. and N.D. did the on-board experiments during the research cruises. F.U.Z. analysed the data from physiology experiments. J.M.P. amplified and sequenced *hupL*, analysed the genome data, did western blots and immunohistochemistry. C.M. and R.A. did the geneFISH. S.H., S.D.W. and P.R.G. did the *in situ* mass spectrometry and analysed the data. V.B. and E.P. did the genome sequencing and assembly. W.B. did the thermodynamic modelling. J.M.P., F.U.Z. and N.D. conceived the study and wrote the paper.

Author Information All *hupL* sequences have been deposited at NCBI under accession numbers FR851255–FR851274. The sequences that make up the genome fragment and the RAST annotation can be found at NCBI under project identification 65421 (accession numbers CAEB01000001–CAEB01000078). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D. (ndubilier@mpi-bremen.de).

Microwave quantum logic gates for trapped ions

C. Ospelkaus¹†, U. Warring¹, Y. Colombe¹, K. R. Brown¹†, J. M. Amini¹†, D. Leibfried¹ & D. J. Wineland¹

Control over physical systems at the quantum level is important in fields as diverse as metrology, information processing, simulation and chemistry. For trapped atomic ions, the quantized motional and internal degrees of freedom can be coherently manipulated with laser light^{1,2}. Similar control is difficult to achieve with radio-frequency or microwave radiation: the essential coupling between internal degrees of freedom and motion requires significant field changes over the extent of the atoms' motion^{2,3}, but such changes are negligible at these frequencies for freely propagating fields. An exception is in the near field of microwave currents in structures smaller than the free-space wavelength^{4,5}, where stronger gradients can be generated. Here we first manipulate coherently (on time-scales of 20 nanoseconds) the internal quantum states of ions held in a microfabricated trap. The controlling magnetic fields are generated by microwave currents in electrodes that are integrated into the trap structure. We also generate entanglement between the internal degrees of freedom of two atoms with a gate operation^{4,6–8} suitable for general quantum computation⁹; the entangled state has a fidelity of 0.76(3), where the uncertainty denotes standard error of the mean. Our approach, which involves integrating the quantum control mechanism into the trapping device in a scalable manner, could be applied to quantum information processing⁴, simulation^{5,10} and spectroscopy^{3,11}.

The quantized mechanical motion and internal states of trapped atoms can be coherently controlled by laser radiation. This optical control has led to the creation of non-classical motional states^{1–3}, multiparticle entanglement¹, scalable quantum logic techniques¹², quantum simulation studies^{13–15} and quantum logic spectroscopy¹¹. The coupling between internal and motional states is provided by the gradient of the driving field over the extent, Δx , of the atoms' motion^{2,3}. For trapped atomic ions in their lowest motional quantum levels, $\Delta x \approx 10$ nm, which is a significant fraction of an optical wavelength, resulting in an efficient coupling between motional and internal states. A microwave field propagating in free space varies only very slightly over Δx , and the corresponding coupling between motional and internal states is highly suppressed.

Notwithstanding the success of laser-based control, it is desirable to develop a microwave approach with equivalent capabilities because of the relative ease of generating and controlling these fields. One method uses a combination of uniform microwave fields and static, state-dependent potentials to provide motional–internal state coupling^{16–19} and individual addressing^{18,20}. Related to this are magnetic microwave or radio-frequency-dressed state potentials for neutral atoms²¹. Here we implement the approach outlined in ref. 4, where the magnetic field from oscillating microwave currents in trap electrodes provides coherent manipulation of the internal state, and the gradient of this field leads to motional–internal state coupling. These interactions suffice for universal quantum information processing, because any unitary operation on quantum bits (qubits) can be decomposed into a sequence of single-qubit rotations and a suitable entangling interaction⁹. By comparison with laser-based approaches, microwave control has the potential to improve significantly the fidelity of operations owing to its reduced sensitivity to 'spectator modes' of ion motion³, its

better control of field amplitudes and phases, and the absence of spontaneous-emission decoherence^{4,16}. Furthermore, microwave near-field control could be incorporated in a chip-level library of transport, junction, storage and microwave manipulation components²² to advance the integration of quantum control in scalable quantum information processing or simulation.

Our apparatus comprises a room-temperature (300 K) surface-electrode ion trap²³ with 10- μm -thick gold electrodes, separated by gaps of 4.5 μm , deposited onto an insulating AlN substrate (Fig. 1). An oscillating potential (amplitude, 25–50 V; frequency, $f_{\text{RF}} = 70.97$ MHz), applied to the radio-frequency electrodes (Fig. 1, RF1 and RF2), provides pseudopotential confinement of $^{25}\text{Mg}^+$ ions in the radial (x and z) directions at a distance of 30 μm from the electrode surface (all other electrodes are held at radio-frequency ground). Along the trap y axis, ions are confined with static potentials applied to control electrodes C1, C3, C4 and C6. The radial field resulting from these potentials is compensated for by suitable potentials applied to electrodes C2 and C5. Single-ion oscillation frequencies along the y direction can be adjusted between 300 kHz and 2 MHz, and radial frequencies can be adjusted between 4 and 10 MHz. Microwave electrodes MW1, MW2 and MW3 support currents, of order 100 mA to 1 A, that produce an oscillating magnetic field, B_{osc} , above the surface and are used to implement quantum control. To minimize thermal effects during these microwave pulses, we chose AlN as the electrode substrate. It provides a strong thermal link between the gold electrodes and a solid copper support through a thin layer of vacuum-compatible epoxy. A static magnetic field, B_0 , parallel to the trap surface and at an angle of 15° with respect to the z axis, provides the internal-state quantization axis. Superimposed circularly (σ^-)-polarized laser beams nearly resonant

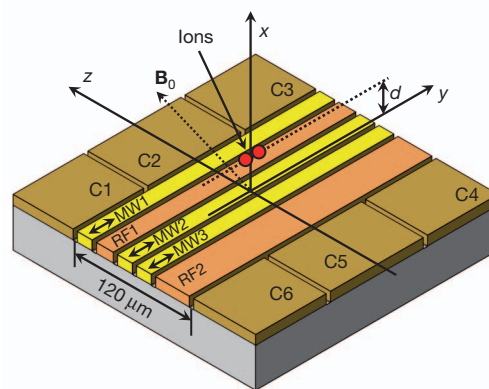


Figure 1 | Central portion of the surface-electrode trap. The control electrodes are labelled C1 to C6, trap radio-frequency electrodes are labelled RF1 and RF2, and microwave lines are labelled MW1 to MW3. The ions are held a distance $d = 30 \mu\text{m}$ above the surface. By comparison with the symmetric trap structure of ref. 4, the asymmetric geometry of the microwave and radio-frequency electrodes used here yields approximately equal microwave currents in all three electrodes for field suppression at the ions while providing a deeper trap.

¹Time and Frequency Division, National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA. †Present addresses: Institute of Quantum Optics, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, and PTB, Bundesallee 100, 38116 Braunschweig, Germany (C.O.); GTRI Georgia Tech, 400 10th Street NW, Atlanta, Georgia 30318, USA (K.R.B., J.M.A.).

with $3^2S_{1/2} \rightarrow 3^2P_{3/2}$ transitions and propagating parallel to \mathbf{B}_0 are used for optical pumping, Doppler laser cooling and state detection through resonance fluorescence. Optical pumping prepares ions in the $|F = 3, m_F = -3\rangle$ hyperfine ground state. Using microwave pulses resonant with hyperfine transitions, we can prepare an arbitrary pure state within the ground-state manifold (Fig. 2a) with fidelity greater than 0.98 (Methods). Furthermore, we can detect the population in any one of these states by applying a series of microwave π -pulses to transfer its population to $|3, -3\rangle$ and then detecting resonance fluorescence on the $|3, -3\rangle \rightarrow |3^2P_{3/2}, F = 4, m_F = -4\rangle$ optical cycling transition.

At a bias field of $B_0 \approx 21.3$ mT, we realize a first-order magnetic-field-independent transition between the qubit states $|3, 1\rangle \equiv |\downarrow\rangle$ and $|2, 1\rangle \equiv |\uparrow\rangle$ with 'carrier' frequency $f_0 \approx 1.69$ GHz (Fig. 2a, solid line). To prepare $|\downarrow\rangle$, after optically pumping the ions to $|3, -3\rangle$, we apply four microwave π -pulses (Fig. 2a, dotted lines). Figure 2b shows Ramsey spectroscopy measurements of the $|\downarrow\rangle \rightarrow |\uparrow\rangle$ transition frequency as a function of magnetic field detuning, δB , from B_0 . At $\delta B = 0$, the transition frequency depends on δB only to second order, a feature that has been shown to allow long coherence times (~ 10 s) despite moderate ambient magnetic field noise²⁴.

Rabi flopping on the $|\downarrow\rangle \rightarrow |\uparrow\rangle$ transition with a complete state exchange time (π -time) of 18.63(3) ns (Fig. 2c) demonstrates the speed of single-qubit operations. This is more than 200 times shorter than a typical π -time (4 μ s) obtained with continuous-wave Raman laser

beams²⁵. Moreover, it is nine orders of magnitude shorter than the coherence time achieved on a similar field-independent transition²⁴. (For very short π -pulses realized with pulsed lasers, see ref. 26.)

To achieve coupling between motional and internal states, we apply sufficiently large oscillating magnetic field gradients at frequencies $f_s \approx f_0 \pm f_r$ near the motional sidebands of the qubit transition carrier frequency, f_0 (where f_r is a radial motional-mode frequency). At the same time, we want to suppress \mathbf{B}_{osc} at the ions because it can cause undesirable off-resonant carrier excitation and a.c. Zeeman shifts, analogous to a.c. Stark shifts for optical fields⁴. In the geometry of Fig. 1, there is a single combination of relative current amplitudes and phases in electrodes MW1, MW2 and MW3 that provides an oscillating magnetic field gradient at the ions without an oscillating magnetic field.

We find this combination by adjusting the amplitudes and phases of the currents to minimize the a.c. Zeeman shifts on the qubit and neighbouring hyperfine transitions. To characterize the resulting field, we map the x - z spatial dependence of the a.c. Zeeman shift on a selected magnetic dipole hyperfine transition by displacing the ion radially from its nominal position with adjustments of the static trap potentials. Within a small region around the oscillating field zero, we expect \mathbf{B}_{osc} to be a quadrupole characterized by a gradient strength, B' , and an angle, α (the angle of one principal axis of the quadrupole with respect to the x axis; see Methods). A fit based on this model and to the experimental a.c. Zeeman shift data yields $B' = 35.3(4) \text{ T m}^{-1}$ and $\alpha = 26.6(7)^\circ$. For comparison, a plane-wave microwave field with the amplitude required to produce the π -time of Fig. 2c (1.9 mT), has a gradient of only 0.068 T m^{-1} . Knowledge of the orientation and gradient of the field can be used to predict sideband Rabi rates and to optimize the overlap with one of the radial motional modes. From the measured gradient and the assumption of optimal alignment of the mode with the gradient, we calculate a sideband π -time⁴ of approximately 190 μ s at a motional frequency of 6.5 MHz for a single ion in its motional ground state (Methods).

After approximately optimizing the field configuration, we implemented motional sideband transitions on one ion and on a pair of ions. As an example of a two-ion sideband application, we perform the following experiments on the selected radial rocking mode. First, two ions are Doppler-cooled and prepared in the $|\uparrow\uparrow\rangle$ state. We then apply a qubit π -pulse followed by a sideband pulse at $f_s = f_0 + f_r$ (here $f_r \approx 6.8$ MHz for the chosen rocking mode). The sideband pulse induces transitions from $|\downarrow\downarrow, n\rangle$ to $|\uparrow\downarrow, n+1\rangle$, $|\downarrow\uparrow, n+1\rangle$ and $|\uparrow\uparrow, n+2\rangle$, where n denotes the motional quantum number²⁷. Populations in $|\downarrow\downarrow\rangle$ are subsequently transferred to $|3, -3\rangle$ and detected by laser fluorescence. In Fig. 3a (blue circles), we show the result of a frequency scan of the sideband pulse. We then repeat the above experiment, but swap the order of the sideband pulse and the carrier π -pulse (Fig. 3a, red squares). In this case, the sideband pulse drives transitions from $|\uparrow\uparrow, n\rangle$ to $|\downarrow\uparrow, n-1\rangle$, $|\uparrow\downarrow, n-1\rangle$ and $|\downarrow\downarrow, n-2\rangle$. If we assume a thermal motional state (valid for Doppler cooling), a fit to the data gives an average occupation number of $\bar{n} = 2.2(1)$.

Multiple sideband cooling cycles can be used to cool a selected motional mode to near the ground state^{2,27}. One sideband cooling cycle consists of transitions from $|\uparrow\uparrow, n\rangle$ to $|\downarrow\uparrow, n-1\rangle$, $|\uparrow\downarrow, n-1\rangle$ and $|\downarrow\downarrow, n-2\rangle$, where the pulse duration (250 μ s) is optimized for maximal sideband asymmetry. This is followed by a repumping laser pulse resonant with the $|\downarrow\rangle \rightarrow |3^2P_{3/2}, m_J = -3/2, m_I = 3/2\rangle$ transition (where m_J and m_I are the components of the total electronic and nuclear angular momenta parallel to the static magnetic field, \mathbf{B}_0). From the relevant Clebsch–Gordan coefficients, we predict a 97% repumping efficiency to $|\uparrow\rangle$ after scattering 20 photons on average. Figure 3b shows a frequency scan of the motion-adding and motion-subtracting sidebands after four sideband cooling cycles. Assuming an approximately thermal distribution after cooling²⁷, we find that $\bar{n} = 0.6(1)$, confirming the sideband cooling. Lower temperatures could not be achieved by adding more cooling cycles, and the final

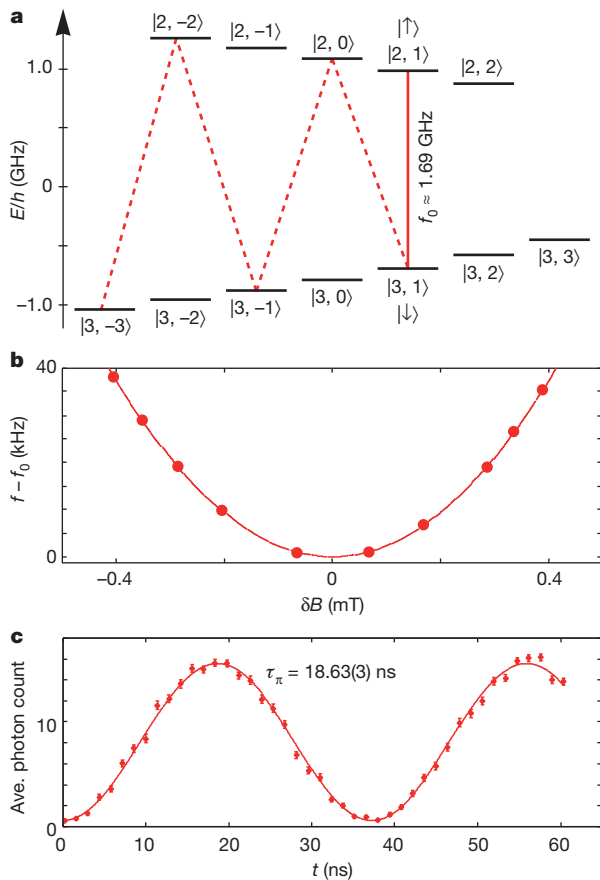


Figure 2 | Level scheme of $^{25}\text{Mg}^+$ (nuclear spin, $I = 5/2$) and spectroscopy of the qubit transition. **a**, Ground-state hyperfine structure of $^{25}\text{Mg}^+$ at $B_0 \approx 21.3$ mT and the microwave transitions used in the experiment. h , Planck's constant. **b**, Frequency of the $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition as a function of bias-field detuning, δB , from the field-independent point, B_0 . The standard errors of the measured frequencies are less than 5 Hz. **c**, Rabi flopping on the $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition induced by microwave current in MW2, the electrode in closest proximity to the ions. Error bars, s.e.m. for the average photon count in 400 μ s. τ_π , π -time.

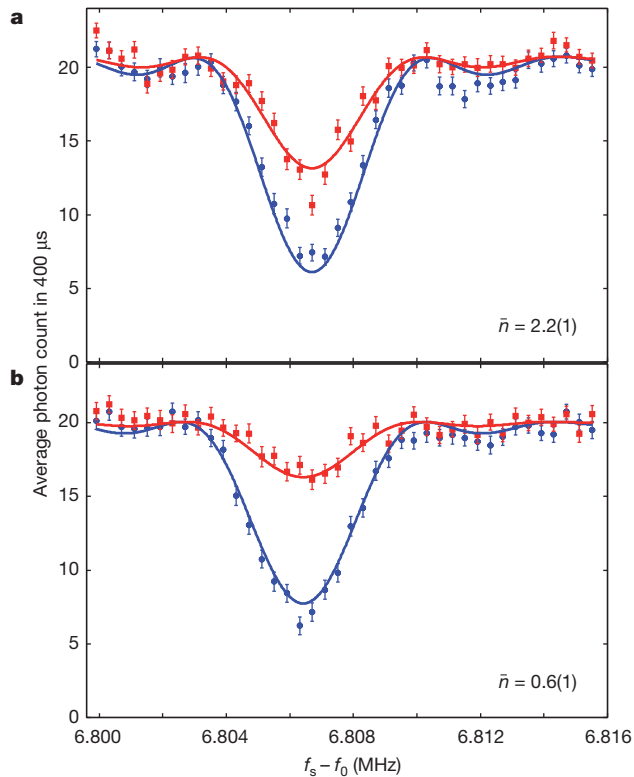


Figure 3 | Microwave motional sideband transitions. **a**, Motion-subtracting $|\uparrow\uparrow, n\rangle \rightarrow \{|\uparrow\downarrow, n-1\rangle, |\downarrow\uparrow, n-1\rangle, |\downarrow\downarrow, n-2\rangle\}$ (red squares) and motion-adding $|\downarrow\downarrow, n\rangle \rightarrow \{|\uparrow\downarrow, n+1\rangle, |\downarrow\uparrow, n+1\rangle, |\uparrow\uparrow, n+2\rangle\}$ (blue circles) sideband transitions for a Doppler-cooled two-ion radial rocking mode. **b**, Sideband transitions after four sideband cooling cycles. Error bars, s.e.m. for the average photon count in 400 μ s.

value of \bar{n} was most probably limited by the repumping photon recoil and by heating from ambient electric field noise²⁷ during the course of the sequence. Separately, we measured the relevant motional heating rate and found it to be as low as $d\bar{n}/dt = 0.2 \text{ ms}^{-1}$ at $f_r = 7.7 \text{ MHz}$; however, higher heating rates were also observed. We did not detect additional heating induced by the microwave pulses.

Simultaneous application of two field gradients at $f_0 \pm (f_r + \delta)$, with $\delta \ll f_r$, is used to implement an entangling two-qubit gate⁴. By comparison with light fields, derivatives of the microwave fields of order higher than linear have negligible effect on the ions⁴, so the ions are only Doppler-cooled and prepared in the $|\downarrow\downarrow\rangle$ state. We then apply the two oscillating field gradients with $f_r = 7.6 \text{ MHz}$ (rocking mode) and $\delta = 4.9 \text{ kHz}$. To suppress decoherence caused by motional frequency instability, we apply the fields in two pulses of duration 200 μ s, reversing the phase of one field in the second pulse²⁸. For properly chosen pulse durations, the internal states ideally end in the entangled state $|\Psi\rangle = (1/\sqrt{2})(|\downarrow\downarrow\rangle - i|\uparrow\uparrow\rangle)$, with the motion restored to the state it was in before the gate operation⁶. Pulse imperfections and decoherence during the operation produce a mixed state with density matrix ρ (after tracing over the motion). We characterize the fidelity of ρ with respect to the ideal outcome, $F_s = \langle\Psi|\rho|\Psi\rangle = (1/2)(\rho_{\uparrow\uparrow,\uparrow\uparrow} + \rho_{\downarrow\downarrow,\downarrow\downarrow}) + |\rho_{\uparrow\uparrow,\downarrow\downarrow}|$, by observing population oscillations as we sweep the phase, ϕ , of a $\pi/2$ -pulse applied to the qubits after creating the entangled state²⁹. The populations $\rho_{\downarrow\downarrow,\downarrow\downarrow}$, $\rho_{\uparrow\downarrow,\uparrow\downarrow} + \rho_{\downarrow\uparrow,\downarrow\uparrow}$ and $\rho_{\uparrow\uparrow,\uparrow\uparrow}$ can be directly determined by fitting histograms for two, one or zero fluorescing ions to the histograms of the total fluorescence signal obtained during the sweep (Methods). The resulting populations are shown in Fig. 4a. The magnitude of the density matrix element $\rho_{\uparrow\uparrow,\downarrow\downarrow}$ is equal to the amplitude, A_{IT} , of the parity $\Pi = (P_{|\uparrow\uparrow\rangle} + P_{|\downarrow\downarrow\rangle}) - (P_{|\uparrow\downarrow\rangle} + P_{|\downarrow\uparrow\rangle})$, which oscillates like $A_{IT}\cos(2\phi + \phi_0)$ (ref. 29; Fig. 4b). By fitting sinusoids to the observed populations, we extract a fidelity of $F_s = 0.76(3)$ with respect to the state $|\Psi\rangle$.

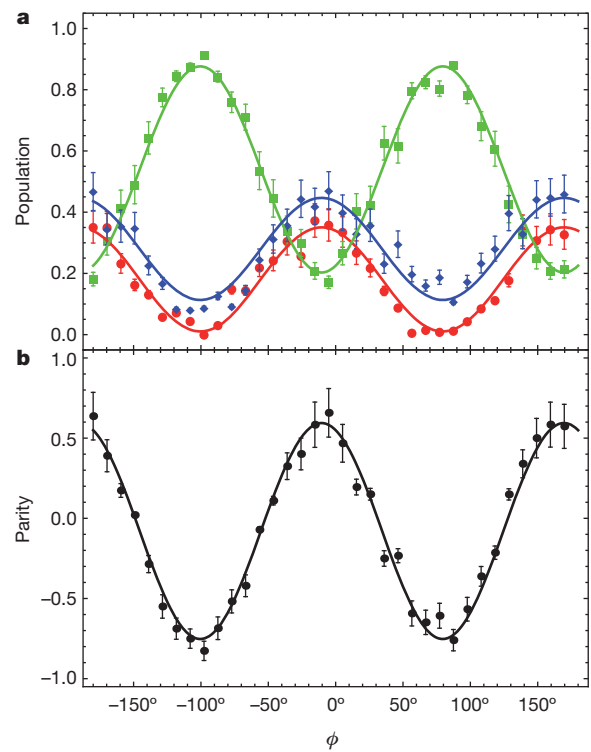


Figure 4 | Populations and parity of the entangled state. After the entanglement operation is applied to the $|\downarrow\downarrow\rangle$ state, an analysis $\pi/2$ -pulse with relative phase ϕ is applied. Ideally, as ϕ is varied, $|\Psi\rangle$ is transformed into superpositions of $|\Psi\rangle$ and $(1/\sqrt{2})(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$, and parity and populations of the two ions oscillate at $\cos(2\phi + \phi_0)$. **a**, Populations in $|\uparrow\uparrow\rangle$ ($P_{|\uparrow\uparrow\rangle}$, red disks), $|\downarrow\downarrow\rangle$ ($P_{|\downarrow\downarrow\rangle}$, blue diamonds) and $|\uparrow\downarrow\rangle$ plus $|\downarrow\uparrow\rangle$ ($P_{|\uparrow\downarrow\rangle} + P_{|\downarrow\uparrow\rangle}$, green squares). **b**, Parity $\Pi = (P_{|\uparrow\uparrow\rangle} + P_{|\downarrow\downarrow\rangle}) - (P_{|\uparrow\downarrow\rangle} + P_{|\downarrow\uparrow\rangle})$. Each data point represents an average of 300 experiments. We determine an entangled-state fidelity of $F_s = 0.76(3)$ from the populations in **a**. Error bars, s.e.m. for the populations and parity.

For scalable implementations of quantum information processing, the quality of all operations demonstrated here, in particular that of the entangling operation, must be improved considerably. This will require significant technical improvements; however, there are no apparent fundamental limits to fidelity. The most important factors currently limiting two-qubit gate fidelity are the stability of motional frequencies in the trap and the precision with which we can suppress \mathbf{B}_{osc} at the positions of the ions. Fluctuations of the motional frequencies are of the order of 1 kHz during the gate, which is a sizeable fraction of the gate detuning, δ , and seem to be primarily caused by time-varying stray potentials on the trap chip. It should be possible to suppress such stray fields by improving the surface quality of the electrodes, by reducing the amount of nearby dielectric material (dust and the residue of materials used during fabrication), by more carefully cleaning the electrode surfaces and by minimizing the exposure of the electrodes to ultraviolet light, which can generate charge through photoemission. We are able to suppress the oscillating field at the positions of the ions to fluctuations with amplitudes of approximately 10 μT , implying an a.c. Zeeman shift of less than 1 kHz. These fluctuations could be further suppressed by improving the amplitude stability and shaping of the microwave pulses and by using more sophisticated microwave current leads.

The presence of many processing zones on the same chip can introduce cross-talk between zones. Cross-talk for two-qubit gates can be suppressed by choosing substantially different mode frequencies for qubits in nearby ‘spectator’ trap zones. For single-qubit rotations, it should be possible to reduce the field to near zero in spectator zones by applying local compensating fields. Composite pulse techniques can also prove useful in enhancing spectator-zone isolation³⁰. Traps smaller

than those used here should lead to faster operations⁴, provided that anomalous motional heating can be suppressed. Multi-ion operations similar to those demonstrated here might also be used for quantum simulation^{5,10}, possibly with less stringent requirements on suppression of cross-talk. For particles lacking the electronic structure needed for laser manipulation and read-out, microwave sidebands would allow excitation of the motion conditioned on the internal-state magnetic moment³¹. The presence of the motional excitation could then be observed with an ancilla ion, either through a shared motional mode¹¹ or through the Coulomb interaction between ions held in separate traps^{3,31–33}. In the second case, this technique might allow a comparison of the proton and antiproton magnetic moments^{3,31}.

METHODS SUMMARY

We describe the oscillating magnetic field geometry in the vicinity of the ions, and the experimental determination of this geometry, by mapping out a.c. Zeeman shifts experienced by the ions as functions of their positions. See the Methods for more details. There we also explain how state populations are extracted from fluorescence measurements to derive the fidelity of the entangled state produced by the two-qubit gate.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 April; accepted 14 June 2011.

- Blatt, R. & Wineland, D. J. Entangled states of trapped atomic ions. *Nature* **453**, 1008–1015 (2008).
- Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. J. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
- Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl Inst. Stand. Technol.* **103**, 259–328 (1998).
- Ospelkaus, C. *et al.* Trapped-ion quantum logic gates based on oscillating magnetic fields. *Phys. Rev. Lett.* **101**, 090502 (2008).
- Chiaverini, J. & Lybarger, W. E. Laserless trapped-ion quantum simulations without spontaneous scattering using microtrap arrays. *Phys. Rev. A* **77**, 022324 (2008).
- Sørensen, A. & Mølmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1974 (1999).
- Solano, E., de Matos Filho, R. L. & Zagury, N. Deterministic Bell states and measurement of the motional state of two trapped ions. *Phys. Rev. A* **59**, R2539–R2543 (1999).
- Milburn, G. J., Schneider, S. & James, D. F. V. Ion trap quantum computing with warm ions. *Fortschr. Phys.* **48**, 801–810 (2000).
- Barenco, A. *et al.* Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).
- Schmied, R., Wesenberg, J. H. & Leibfried, D. Optimal surface-electrode trap lattices for quantum simulation with trapped ions. *Phys. Rev. Lett.* **102**, 233002 (2009).
- Schmidt, P. O. *et al.* Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
- Knoop, M., Hilico, L. & Eschner, J. (eds) *Modern Applications of Trapped Ions* (J. Phys. B Vol. 42, Institute of Physics, 2009).
- Porras, D. & Cirac, J. I. Effective quantum spin systems with trapped ions. *Phys. Rev. Lett.* **92**, 207901 (2004).
- Friedenauer, A., Schmitz, H., Glueckert, J. T., Porras, D. & Schaetz, T. Simulating a quantum magnet with trapped ions. *Nature Phys.* **4**, 757–761 (2008).
- Kim, K. *et al.* Quantum simulation of frustrated Ising spins with trapped ions. *Nature* **465**, 590–593 (2010).
- Mintert, F. & Wunderlich, C. Ion-trap quantum logic using long-wavelength radiation. *Phys. Rev. Lett.* **87**, 257904 (2001).
- Ciaramicoli, G., Marzoli, I. & Tombesi, P. Scalable quantum processor with trapped electrons. *Phys. Rev. Lett.* **91**, 017901 (2003).
- Johanning, M. *et al.* Individual addressing of trapped ions and coupling of motional and spin states using RF radiation. *Phys. Rev. Lett.* **102**, 073004 (2009).
- Förster, L. *et al.* Microwave control of atomic motion in optical lattices. *Phys. Rev. Lett.* **103**, 233001 (2009).
- Wang, S. X., Labaziewicz, J., Ge, Y., Shewmon, R. & Chuang, I. L. Individual addressing of ions using magnetic field gradients in a surface-electrode ion trap. *Appl. Phys. Lett.* **94**, 094103 (2009).
- Fortágh, J. & Zimmermann, C. Magnetic microtraps for ultracold atoms. *Rev. Mod. Phys.* **79**, 235–289 (2007).
- Amini, J. M. *et al.* Toward scalable ion traps for quantum information processing. *New J. Phys.* **12**, 033031 (2010).
- Seidelin, S. *et al.* Microfabricated surface-electrode ion trap for scalable quantum information processing. *Phys. Rev. Lett.* **96**, 253003 (2006).
- Langer, C. *et al.* Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).
- Jost, J. D. *et al.* Entangled mechanical oscillators. *Nature* **459**, 683–685 (2009).
- Campbell, W. C. *et al.* Ultrafast gates for single atomic qubits. *Phys. Rev. Lett.* **105**, 090502 (2010).
- King, B. E. *et al.* Cooling the collective motion of trapped ions to initialize a quantum register. *Phys. Rev. Lett.* **81**, 1525–1528 (1998).
- Hayes, D. *et al.* Coherent error suppression in spin-dependent force quantum gates. Preprint at (<http://arxiv.org/abs/1104.1347>) (2011).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Levitt, M. H. Composite pulses. *Prog. Nucl. Magn. Reson. Spectrosc.* **18**, 61–122 (1986).
- Heinzen, D. J. & Wineland, D. J. Quantum-limited cooling and detection of radio-frequency oscillations by laser-cooled ions. *Phys. Rev. A* **42**, 2977–2994 (1990).
- Brown, K. R. *et al.* Coupled quantized mechanical oscillators. *Nature* **471**, 196–199 (2011).
- Harlander, M., Lechner, R., Brownnutt, M., Blatt, R. & Hänsel, W. Trapped-ion antennae for the transmission of quantum information. *Nature* **471**, 200–203 (2011).

Acknowledgements We thank M. J. Biercuk, J. J. Bollinger and A. P. VanDevender for experimental assistance, J. C. Bergquist, C. W. Chou and T. Rosenband for the loan of a fibre laser, R. Jördens and E. Knill for comments on the manuscript, and D. Hanneke and J. P. Home for discussions. We thank P. Treutlein for discussions on microfabrication techniques. This work was supported by IARPA, the ONR, DARPA, the NSA, Sandia National Laboratories and the NIST Quantum Information Program. This paper, a submission of NIST, is not subject to US copyright.

Author Contributions C.O. participated in the design of the experiment and built the experimental apparatus, collected data, analysed results and wrote the manuscript. U.W. participated in building the experimental apparatus, collected data and analysed results. Y.C. developed chip fabrication methods and fabricated the ion trap chip. K.R.B. participated in the design of the experiment, developed chip fabrication methods and helped build parts of the experiment. J.M.A. developed chip fabrication methods and automated experiment control and data taking. D.L. participated in the design of the experiment, collected data and maintained the laser systems. D.J.W. participated in the design and analysis of the experiment. All authors discussed the results and the text of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.O. (christian.ospelkaus@iqo.uni-hannover.de).

METHODS

Determination of the local microwave field. Within a small region $\{\delta x, \delta y, \delta z\}$ around the oscillating microwave field null point, we expect \mathbf{B}_{osc} to be a quadrupole

$$\mathbf{B}_{\text{osc}} = B' \cos(2\pi f_s t) \begin{pmatrix} \cos(2\alpha) & 0 & \sin(2\alpha) \\ 0 & 0 & 0 \\ \sin(2\alpha) & 0 & -\cos(2\alpha) \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \\ \delta z \end{pmatrix}$$

in the coordinate system of Fig. 1, where B' and α (the angle of one principal axis of the quadrupole with respect to the x direction in Fig. 1) characterize the strength and orientation of the quadrupole, respectively. We map the x - z spatial dependence of the a.c. Zeeman shift on the magnetic dipole transition $|3, 1\rangle \rightarrow |2, 0\rangle$ by displacing the ion radially from its nominal position with adjustments of the control electrode potentials that are derived from simulations. We determine the a.c. Zeeman shifts for a set of radial displacements $\{\delta x, \delta z\}$ and find that $B' = 35.3(4) \text{ T m}^{-1}$ and $\alpha = 26.6(7)^\circ$ from a fit to this data. As a consistency check, we probe the relevant one-ion sideband with frequency near 6.5 MHz and find a π -time for the ion in its motional ground state of approximately 260 μs (extracted by a model fit). The discrepancy of 27% between the observed π -time and that inferred from the measured gradient is most probably due to a misalignment between the maximal gradient direction and the axis of the interrogated motional mode. (To maintain a reasonable trap depth, the motional mode could not be perfectly overlapped with the maximal gradient.)

Determination of populations from state-dependent fluorescence. During one detection period (duration, 100 μs), we typically detect ~ 0.3 counts on average if both ions are projected into $|\uparrow\rangle$, and approximately 11 additional counts for each ion in state $|\downarrow\rangle$.

Reference histograms are derived by recording photon counts for an on-resonance Ramsey experiment with two ions where the phase, θ , of the second $\pi/2$ -pulse was varied. We establish in independent experiments that the net effect of transferring in and out of the qubit basis, and the Ramsey pulses, gives population errors not exceeding 0.02. We therefore assume the respective probabilities of observing zero, one or two fluorescing ions to be

$$P_0(\theta) = \cos^4(\theta/2)$$

$$P_1(\theta) = \sin^2(\theta)/2$$

$$P_2(\theta) = \sin^4(\theta/2)$$

We model each histogram for two, one or zero fluorescing ions as a sum of three Poisson distributions with variable weights and average counts. The weights and averages are determined by a simultaneous least-squares fit to the experimentally determined histograms for all phases θ . To reduce the number of free parameters, we fix the two-ion average to be twice the one-ion average. With this choice, the fit contains eight free parameters and has a reduced χ^2 value of 1.34. Using the histograms representing two, one and zero fluorescing ions, we do a second least-squares fit to the histograms obtained while scanning the phase, ϕ , of the $\pi/2$ -pulse following the entangling gate. Each population is fitted to functions $a_k \cos(2\phi + \phi_0) + a_{0,k}$, where a_k and $a_{0,k}$ ($k = 0, 1$ or 2 fluorescing ions) are pairs of constants and ϕ_0 is a common phase offset. The least-squares fits are shown in Fig. 4a. These fits directly yield the parity amplitude, A_{II} , the phase offset, ϕ_0 , and the offset from zero shown in Fig. 4b.

The final uncertainty in the fidelity, F_s , is derived from the uncertainties in the population fits. In addition to the data analysis described above, we apply several other parameterizations of the histograms to fit to the Ramsey reference experiment. In all instances, the resulting gate fidelities agree with our stated result to within the respective uncertainties.

Quantum gates and memory using microwave-dressed states

N. Timoney¹, I. Baumgart¹, M. Johanning¹, A. F. Varón¹, M. B. Plenio², A. Retzker² & Ch. Wunderlich¹

Trapped atomic ions have been used successfully to demonstrate¹ basic elements of universal quantum information processing. Nevertheless, scaling up such methods to achieve large-scale, universal quantum information processing (or more specialized quantum simulations^{2–5}) remains challenging. The use of easily controllable and stable microwave sources, rather than complex laser systems^{6,7}, could remove obstacles to scalability. However, the microwave approach has drawbacks: it involves the use of magnetic-field-sensitive states, which shorten coherence times considerably, and requires large, stable magnetic field gradients. Here we show how to overcome both problems by using stationary atomic quantum states as qubits that are induced by microwave fields (that is, by dressing magnetic-field-sensitive states with microwave fields). This permits fast quantum logic, even in the presence of a small (effective) Lamb–Dicke parameter (and, therefore, moderate magnetic field gradients). We experimentally demonstrate the basic building blocks of this scheme, showing that the dressed states are long lived and that coherence times are increased by more than two orders of magnitude relative to those of bare magnetic-field-sensitive states. This improves the prospects of microwave-driven ion trap quantum information processing, and offers a route to extending coherence times in all systems that suffer from magnetic noise, such as neutral atoms, nitrogen-vacancy centres, quantum dots or circuit quantum electrodynamic systems.

Using laser light for coherent manipulation of quantum bits (qubits) gives rise to fundamental issues, notably unavoidable spontaneous emission that destroys quantum coherence^{8,9}. The difficulty of cooling a collection of ions to their motional ground state and the long duration of this process in the presence of spurious heating of Coulomb crystals limits the fidelity of quantum logic operations in laser-based quantum gates and thus hampers scalability. This limitation is only partly removed by the use of ‘hot’ gates^{10,11}. Technical challenges in accurately controlling the frequency and intensity of laser light as well as in targeting a large number of laser beams of high intensity on trapped ions are further obstacles to scalability.

These issues associated with the use of laser light for scalable QIP have led to the development of novel ways of performing conditional quantum dynamics with trapped ions that rely on radio-frequency or microwave radiation instead of laser light^{6,7,12–15}. Radio-frequency or microwave radiation can be used in quantum gates through the use of magnetic-gradient-induced coupling between spin states of ions¹⁶, thus averting the technical and fundamental issues of scalability that were described above. Furthermore, the sensitivity to motional excitation of ions is reduced in such schemes. A drawback of magnetic-gradient-induced coupling is the necessity of using magnetic-field-sensitive states for conditional quantum dynamics, thus making qubits susceptible to ambient field noise and shortening their coherence times. This issue is shared with some optical ion trap schemes for QIP that usually rely on magnetic-field-sensitive states for conditional quantum dynamics, such as geometric gates¹, limiting the coherence time of qubit states typically to a few milliseconds. In an effort to extend the coherence times of atomic states, two-qubit entangled states

forming a decoherence-free subspace have been created^{17,18}. Recently, transfer between field-sensitive states used for conditional quantum dynamics and field-insensitive states used for storage of quantum information has been utilized¹⁹.

The relevant noise source in this case, namely magnetic field fluctuations, is not featureless white noise but tends to have a limited bandwidth. In this context, techniques were proposed for prolonging coherence times by subjecting the system to a rapid succession of pulses leading to a decoupling from the environment. This technique, termed bang–bang control²⁰, and its continuous version²¹ can be applied usefully in a variety of systems including hybrid atomic physics and nanophysics technologies. Recent work includes the experimental demonstration of optimized pulse sequences made for suppression of qubit decoherence (see refs 22, 23 and references therein).

Here we encode qubits in microwave-dressed states requiring only continuous, constant-intensity microwave fields. This scheme protects qubits from magnetic field fluctuations and, importantly, simultaneously allows fast quantum logic even for small Lamb–Dicke parameters and, therefore, moderate magnetic field gradients. Microwave-generating elements for the coherent manipulation of qubits can be integrated into microstructured ion traps⁷ such that QIP can be realized using scalable ion chips. Thus, this novel scheme is a significant step towards the integration of elements required for QIP on a scalable ion chip. Moreover, the ideas presented here are generic and can be applied to all laser- or microwave-based QIP.

We theoretically describe the scheme for storage and single- and multiqubit quantum gate operation, and then present experimental demonstrations of storage and information processing of quantum information that demonstrate gains of more than two orders of magnitude in coherence times over approaches involving bare states.

We consider a typical energy-level configuration (Fig. 1a). Encoding quantum information in the subspace spanned by two magnetically sensitive $m_F = \pm 1$ states, $|-1\rangle$ and $|+1\rangle$, will lead to a rapid loss of coherence owing to fluctuating magnetic fields. We first note that microwave-dressed states create a subspace spanned by the two states $|D\rangle = (|-1\rangle - |+1\rangle)/\sqrt{2}$ (a dark state) and $|0'\rangle$ ($m_F = 0$) of the same manifold. This subspace is separated from the other eigenstates of the system, $|u\rangle$ and $|d\rangle$, by a finite energy gap (the dressing fields’ Rabi frequency is denoted by Ω , and $|\Delta_- - \Delta_+| \ll \Omega$ should hold for the detunings Δ_- and Δ_+ ; refs 24, 25). The energetically degenerate states $|D\rangle$ and $|0'\rangle$ are not coupled by magnetic field fluctuations. As a consequence of the energy gap between these states and the states $|u\rangle$ and $|d\rangle$, a relative phase change between $|-1\rangle$ and $|+1\rangle$ acquires an energy penalty and dephasing is strongly suppressed as long as the spectral power density of the magnetic field fluctuations at the frequency corresponding to the energy gap is negligible. We note that the levels $|D\rangle$ and $|0'\rangle$ are energetically degenerate independent of the applied microwave Rabi frequency, Ω , and are hence stable against its power fluctuations. Therefore, we consider the qubit encoded in the subspace spanned by $|D\rangle$ and $|0'\rangle$. The microwave fields do not couple any states in the qubit subspace and thus do not limit the phase coherence.

¹Faculty of Science and Technology, Department of Physics, University of Siegen, 57068 Siegen, Germany. ²Institute for Theoretical Physics, University of Ulm, 89069 Ulm, Germany.

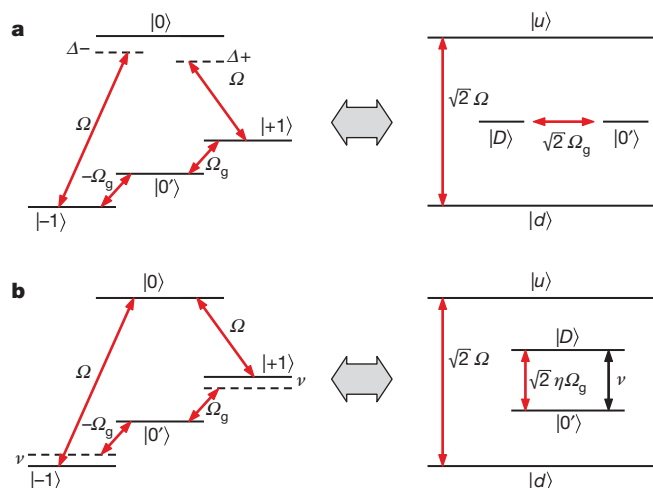


Figure 1 | Microwave-dressed qubit states. Left, atomic states and microwave-dressing fields with Rabi frequency Ω . $| -1 \rangle$ and $| +1 \rangle$ represent magnetic-field-sensitive states, and have, for instance, $m_F = \pm 1$, and $| 0 \rangle$ and $| 0' \rangle$ represent magnetically insensitive levels (and are, for instance, the $m_F = 0$ states of two different hyperfine manifolds). Right, dressed states. $| D \rangle = (| -1 \rangle - | +1 \rangle) / \sqrt{2}$ and $| 0' \rangle$ are separated by an energy gap from two other states, $| u \rangle = (| B \rangle + | 0 \rangle) / \sqrt{2}$ and $| d \rangle = (| B \rangle - | 0 \rangle) / \sqrt{2}$, where $| B \rangle = (| -1 \rangle + | +1 \rangle) / \sqrt{2}$. Thus, a qubit with states $| D \rangle$ and $| 0' \rangle$ is created that is resilient to magnetic field fluctuations and to microwave power fluctuations. **a**, A radio-frequency field with Rabi frequency Ω_g is used to implement general single-qubit quantum gates. **b**, Conditional quantum dynamics coupling electronic degrees of freedom to the motion of trapped atoms is achieved when the radio-frequency field is detuned from the qubit's carrier transition frequency by the vibrational frequency, ν .

The nature of the protected subspace depends on the relative phase between the microwave fields on the transitions $| 0 \rangle \leftrightarrow | \pm 1 \rangle$. The above description applies for a relative phase of zero, whereas for a relative phase of π the protected subspace is spanned by $| B \rangle = (| -1 \rangle + | +1 \rangle) / \sqrt{2}$ and $| 0' \rangle$ (Methods). We considered both cases in the experiments reported below.

The application of one or two additional radio-frequency fields with Rabi frequency Ω_g , possible detuning Δ_g and a possible initial relative phase of π (Methods) then permits the implementation of general single-qubit rotations in the protected subspace spanned by $| B \rangle$ and $| 0' \rangle$ in the case of one additional field and by $| D \rangle$ and $| 0' \rangle$ in the case of two additional fields. Choosing $\Delta_g = 0$ yields the Hamiltonian $H = \Omega_g (| D \rangle \langle 0' | + | 0' \rangle \langle D |)$, which in a Bloch-sphere picture describes arbitrary rotations about the x axis in the protected space. A rotation about the z axis is obtained for $\Delta_g \neq 0$ and $\Omega_g \neq 0$ under the conditions $\Omega_g \ll \Delta_g$ and $\nu \ll \Delta_g$, where ν is the vibrational mode frequency of the harmonically trapped ions' motion. The specific case in which $\Delta_g = \nu$, that is, tuning to the motional sideband, will be discussed later as it corresponds to coupling of the electronic degree of freedom to the motional degree of freedom.

Hyperfine states of electrostatically trapped $^{171}\text{Yb}^+$ ions characterized by quantum numbers $F = 0$ and $F = 1$ in the electronic ground state $S_{1/2}$ are used in the experiments reported here (Methods).

After the initial preparation of an ion in the $| 0 \rangle$ ($F = 0$) state by optical pumping, the procedure to generate and detect dressed states can be thought of as being divided into three segments (Fig. 2, inset). First (up to time $t = T_1$), an incomplete stimulated Raman adiabatic passage²⁴ (STIRAP) is used to transfer the system adiabatically from the atomic-state basis to the dressed-state basis. Second (up to $t = T_2$), the amplitudes of the dressing fields are kept constant for a holding time, $T = T_2 - T_1$, because we are interested in creating and using dressed states (as opposed to population transfer). Between $t = T_1$ and $t = T_2$, dressed states are present (Fig. 1a, b, right-hand side) and quantum operations with them are implemented through application of radio-frequency fields. For a single-qubit gate between dressed

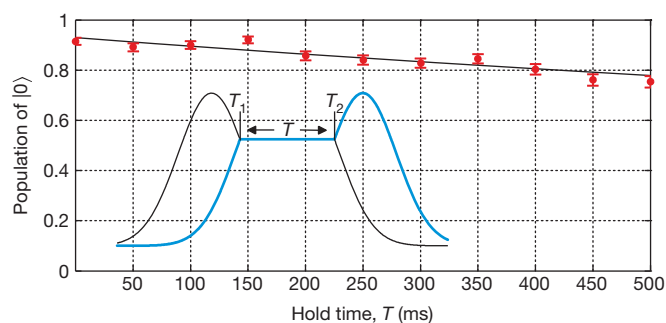


Figure 2 | Lifetime of the dressed state $| D \rangle$. To prepare the dressed state $| D \rangle$, the dressing fields' amplitudes are ramped up adiabatically until time T_1 . Then these amplitudes are kept constant until time T_2 . Between T_1 and T_2 , the dressed states shown in Fig. 1 (right-hand side) are present. Finally, after time T_2 , the STIRAP sequence is completed, adiabatically transferring the dressed state $| D \rangle$ back to a bare ionic state. The hold time is $T = T_2 - T_1$. If state $| D \rangle$ stays invariant during time T , then the final ionic state will be $| 0 \rangle$. The data points indicate the probability to find the ion in state $| 0 \rangle$ upon completion of this sequence as a function of T . The solid line represents a fitted exponential giving a lifetime of $1,700 \pm 300$ ms. Number of repetitions, $n = 300$; error bars, s.d. Experimental parameters are given in Supplementary Information. Inset, qualitative time dependence of the two dressing fields' amplitudes (black and blue; set on resonance with the atomic transitions $| -1 \rangle \leftrightarrow | 0 \rangle$ and $| +1 \rangle \leftrightarrow | 0 \rangle$).

states (Fig. 1a, right-hand side) the radio-frequency field is on resonance with the $| 0' \rangle \leftrightarrow | \pm 1 \rangle$ transition, and for a multiqubit gate (Fig. 1b) the radio frequency is detuned from this transition by the vibrational mode frequency, ν . Third ($t > T_2$), the STIRAP sequence is completed when the ion is transferred from the dressed-state basis back to the atomic states. Any dephasing of a dressed state or transitions to other states during the holding time, T , gives rise to imperfect population transfer during step three of this sequence that returns the system to the atomic-state basis. This is described in more detail in Methods for the creation and detection of the state $| D \rangle$.

In Fig. 2, we show the effectiveness of preparation and detection of $| D \rangle$ as a function of the holding time, T . The measurement shown there extends over 500 ms and an exponential fit of the data yields a lifetime for this state of $1,700 \pm 300$ ms. This lifetime is limited by magnetic field fluctuations that couple $| D \rangle$ to other states and the creation of a different dark state through phase fluctuations between the two microwave fields. This represents a remarkable improvement, of more than two orders of magnitude, in the dependence on the magnetic fluctuations, as the dephasing time of the magnetically sensitive bare atomic states $| \pm 1 \rangle$ in this apparatus has a measured coherence time not exceeding 5.3 ms.

We have investigated the effectiveness of the STIRAP process for creating dressed states and transferring them into the final state, as a function of the parameters characterizing the pulse sequence, and have found this technique to be robust over a wide range of experimental parameters (Methods).

To demonstrate the enhanced coherence time of this qubit, we have conducted Rabi- and Ramsey-type measurements. First, the system is transferred to the dressed-state basis and then an additional radio-frequency field with Rabi frequency Ω_g is applied that induces Rabi oscillations between the dressed states $| B \rangle$ and $| 0' \rangle$. Using the state $| B \rangle$ has the advantage that no phase of π between the radio-frequency fields is needed (compare with Fig. 1). Completing the STIRAP cycle then maps the system to an atomic state that depends on the position of the atom in the dressed-state Rabi cycle. Results are shown in Fig. 3a. The Rabi oscillations are sustained over 550 ms, demonstrating the long-lived coherence of the dressed states when driven by radio-frequency radiation.

These experiments also demonstrate that coherent transfer to the dressed-state basis and subsequent application of a Rabi pulse prepares a coherent superposition in the qubit space that can then be read out

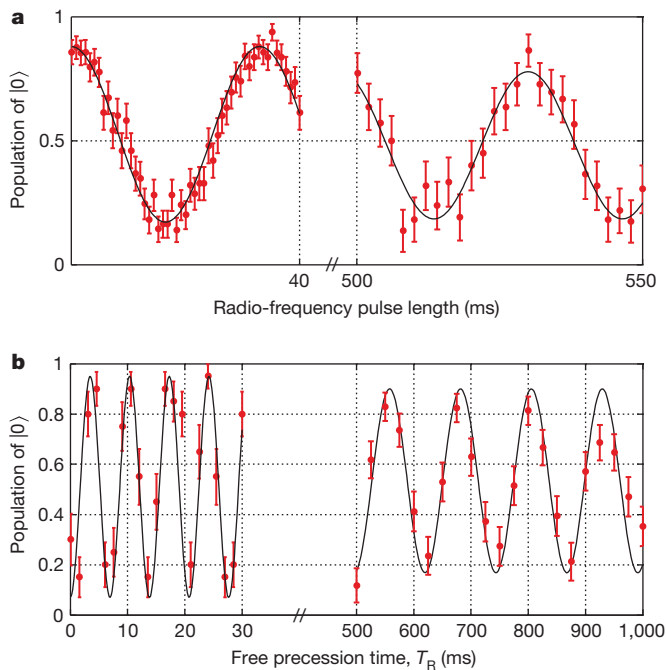


Figure 3 | Single-qubit gates with dressed states. **a**, Rabi oscillations between dressed state $|B\rangle$ and $|0'\rangle$, induced by a radio-frequency field for times of up to more than 500 ms. The population in $|B\rangle$ is mapped onto state $|0\rangle$ (data points) at the end of the STIRAP and detection sequence. $n = 50$ (times up to 40 ms) and $n = 25$ (times more than 500 ms); error bars, s.d. **b**, Ramsey-type measurement preparing a coherent superposition of $|B\rangle$ and $|0'\rangle$ and probing it after time T_R , the free precession time. Two radio-frequency $\pi/2$ -pulses separated by a period T_R of free evolution are applied to the qubit transition. The radio frequency is slightly detuned from resonance (near 10.265 MHz, corresponding to the $|-1\rangle \leftrightarrow |0'\rangle$ and $|+1\rangle \leftrightarrow |0'\rangle$ transitions) yielding Ramsey fringes with period $1/(144.4 \text{ Hz})$ between 0.1 and 30 ms of free precession time. For the measurement between 500 and 1,000 ms, the period is $1/(8.069 \text{ Hz})$ owing to a change in the detuning of the radio-frequency field. $n = 20$ (times up to 30 ms) and $n = 40$ (times 500–1,000 ms); error bars, s.d. Experimental parameters are given in Supplementary Information.

efficiently after completion of the STIRAP cycle. This forms the basis for the measurements of the lifetime of coherent superpositions in the protected subspace, which we now turn to.

Our Ramsey-type experiments test the dephasing time of the dressed-state qubit. Figure 3b shows that coherence is preserved for more than 1,000 ms, close to the absolute limit, of about 1,700 ms, set by the lifetime of the state $|D\rangle$ and more than two orders of magnitude longer than the dephasing time of the atomic states $|-1\rangle$ and $|+1\rangle$, making this scheme ideal for realizing a quantum memory.

Multiqubit gates can be realized by coupling the electronic qubit to the motion of the ions^{10,11,26}. Several schemes for achieving such conditional quantum dynamics are possible using dressed states. Here we outline the scheme illustrated in Fig. 1b. We use a pair of radio-frequency fields on the $|0'\rangle \leftrightarrow |\pm 1\rangle$ transitions (or a single radio-frequency field, depending on the choice of the phase between the dressing fields; Methods). These radio-frequency fields are detuned by ν from resonance and are thus in resonance with the first motional sideband (Fig. 1b). In the dressed-state basis, this couples the $|B\rangle \leftrightarrow |0'\rangle$ (one additional field) or $|D\rangle \leftrightarrow |0'\rangle$ (two additional fields) qubit resonance of the protected subspace to the vibrational mode of the ion string. The coupling of $|D\rangle \leftrightarrow |0'\rangle$ or $|B\rangle \leftrightarrow |0'\rangle$, as appropriate, at zeroth order in the Lamb–Dicke parameter (carrier transition) will be cancelled in the rotating-wave approximation owing to the higher energy (by an amount corresponding to Ω) of $|B\rangle$ or $|D\rangle$, and $|D\rangle$ or, respectively, $|B\rangle$ will be coupled only at first order in the Lamb–Dicke parameter such that we obtain a Hamiltonian of the form (Methods)

$$H = \sqrt{2}\eta\Omega_g(|D\rangle\langle 0'|e^{i\delta t} + \text{h.c.})(b^+ - b)$$

where η is the Lamb–Dicke parameter, b and b^+ are respectively the annihilation and creation operators of the trap harmonic oscillator († denotes adjoint), and h.c. denotes Hermitian conjugate. Importantly, because this gate has no carrier part, a small η value can be compensated for by increasing the radio-frequency power while obeying $\eta\Omega_g \ll \nu$. This will allow moderate static^{5,6,12} or oscillating⁷ magnetic field gradients to be used when realizing multiqubit quantum gates with radio-frequency or microwave radiation. Also, this scheme permits the realization of other types of gate, including gates with ions in thermal motion^{10,11}.

We now outline how the dressed-state scheme for QIP can be applied to a collection of trapped ions. The long wavelength of the microwave radiation requires the use of a static magnetic field gradient, if individual addressing in frequency space is desired. A gradient of a static or oscillating magnetic field is required for the generation of the coupling of the electronic degrees of freedom to the motional degrees of freedom when microwave or radio-frequency radiation is used for quantum gates. Therefore, this does not represent an additional experimental requirement.

In the absence of a static magnetic field gradient, a single pair of microwave-dressing fields is sufficient to dress all ions. In the presence of a static gradient, each ion would require its own pair of driving fields. This can be accomplished efficiently by using a microwave frequency comb where the frequency spacing coincides with the change in static Zeemann shift between neighbouring ions (Fig. 4). Thus, the dressed-state structure for each ion remains essentially the same as in the single-ion case resulting in robust memory. For both single-qubit gates and multiqubit gates, individual addressing is achieved by choosing the frequency of the radio-frequency field. Corrections are small under the condition that $\Omega_g \ll \Delta E_Z/\hbar$, where ΔE_Z is the difference in Zeeman shifts between two neighbouring ions and \hbar is Planck's constant divided by 2π . An important advantage over the standard quantum computing scheme is that the carrier frequency and the sideband can be chosen by interference and not by setting the frequency on resonance with one of these transitions; that is, the choice is set by the phase difference between the two radio-frequency driving fields (Methods). This allows for fast multiqubit gates that are not limited by $\Omega_g \ll \nu$ but by $\eta\Omega_g \ll \nu$; that is, such gates are possible even in the presence of a small (effective) Lamb–Dicke parameter.

The insight gained in this work removes a major obstacle to laser-free QIP (but is also applicable to laser-based schemes). Furthermore, this scheme is not restricted to trapped ions and is in fact applicable to

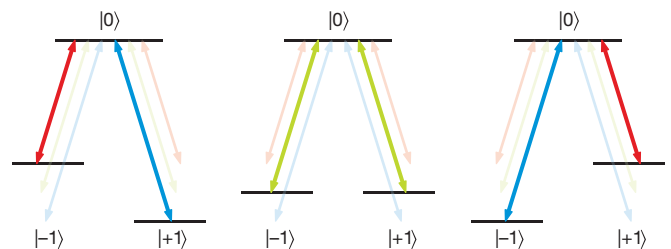


Figure 4 | Multiqubit scheme for three ions exposed to a magnetic field gradient. The magnetic field at one of the two outer ions has the same magnitude but is opposite in direction to the field at the other outer ion, producing opposite, symmetrical Zeemann energy shifts of the levels $|-1\rangle$ and $|+1\rangle$. The arrows represent three pairs of microwave-dressing fields (each pair indicated by a different colour) from a comb with a frequency interval corresponding to the Zeemann shift between neighbouring ions. For each off-resonance microwave field with detuning Δ acting on a particular ion, there is a second field with equal and opposite detuning, $-\Delta$. Hence, each such off-resonance pair will couple only to either $|B\rangle$ or $|D\rangle$ (depending on the relative phase between the microwave fields) leading to equal and opposite a.c. Zeemann shifts, which then cancel.

other physical systems in which dephasing due to external perturbations has a role, for instance neutral atoms²⁷ and solid-state systems such as nitrogen-vacancy centres in diamond²⁸, ion-doped crystals²⁸, and circuit quantum electrodynamics²⁹.

METHODS SUMMARY

The experimental system. The atomic hyperfine states of the electronic ground state of the $^{171}\text{Yb}^+$ ions used in this work can be mapped to the general scheme shown in Fig. 1 as follows: $|F=0\rangle \leftrightarrow |0\rangle$, $|F=1, m_F=-1\rangle \leftrightarrow |-1\rangle$, $|F=1, m_F=+1\rangle \leftrightarrow |+1\rangle$, $|F=1, m_F=0\rangle \leftrightarrow |0'\rangle$ (ref. 30).

The effectiveness of the STIRAP process for creating dressed states and transferring them into the final state has been investigated in detail as a function of the parameters characterizing the pulse sequence (Methods).

Single-qubit gate. We refer to $\{|0\rangle, |D\rangle\}$ as the D qubit and to $\{|0'\rangle, |B\rangle\}$ as the B qubit in what follows. The Hamiltonian reads:

$$\begin{aligned} H_{\text{sqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_{-1}t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_1t} + \text{h.c.}) \\ & + \Omega_g \cos(\lambda_0 t)(|-1\rangle\langle 0'| + \text{h.c.}) \\ & + \Omega_g \cos(\lambda_0 t + \phi)(|1\rangle\langle 0'| + \text{h.c.}) \end{aligned}$$

Here, ω_0 is the zero-field hyperfine splitting, $\omega_{\pm 1}$ are the microwave frequencies with their initial phase difference θ , λ_0 is the radio frequency, and ϕ is the initial phase difference between the radio frequency fields. For $\theta = 0$ and $\phi = \pi$ we obtain the D qubit, and for $\theta = \pi$ and $\phi = 0$ we obtain the B qubit. If we set $\omega_{-1} = \omega_0 + \lambda_0$ and $\omega_1 = \omega_0 - \lambda_0$, then by moving to the interaction picture with respect to the time-independent part we find in the rotating wave approximation (RWA), for the D qubit, that:

$$\begin{aligned} H_{\text{sqg}}^I = & \sqrt{2}\Omega(|B\rangle\langle 0| + \text{h.c.}) + \sqrt{2}\Omega_g(|D\rangle\langle 0'| + \text{h.c.}) \\ = & \frac{\Omega}{\sqrt{2}}|u\rangle\langle u| - \frac{\Omega}{\sqrt{2}}|d\rangle\langle d| + \sqrt{2}\Omega_g(|D\rangle\langle 0'| + \text{h.c.}) \end{aligned}$$

The first two terms shift states $|u\rangle$ and $|d\rangle$ (which both contain $|B\rangle$) away from $|D\rangle$, and the second part creates the gate. The interactions created by the two microwave fields with Rabi frequency Ω and the radio-frequency fields of frequency Ω_g (Fig. 1a) permit the implementation of this single-qubit gate. The two radio-frequency fields have a phase difference of π at the initial time of the pulse, which realizes the coupling to $|D\rangle$ and not to $|B\rangle$.

Alternatively, we may work with the B qubit (Methods).

Multiqubit gate. The levels and fields of Fig. 1b (for the D qubit) are described by the Hamiltonian

$$\begin{aligned} H_{\text{mqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_{-1}t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_1t} + \text{h.c.}) \\ & + \Omega_g \cos[(\lambda_0 - \delta)t](|-1\rangle\langle 0'| + \text{h.c.}) \\ & + \Omega_g \cos[(\lambda_0 - \delta)t + \theta + \pi](|1\rangle\langle 0'| + \text{h.c.}) \\ & + vb^+b|1\rangle\langle 1| + \lambda(|-1\rangle\langle -1| - |1\rangle\langle 1|)(b+b^+) \end{aligned}$$

where λ is proportional to the magnetic gradient as in ref. 6, δ is the detuning of the radio frequency field, $\theta = 0$ for the D qubit and $\theta = \pi$ for the B qubit.

In the interaction picture with respect to the microwaves in RWA, and after the polaron transformation $U = e^{\eta|+1\rangle\langle +1| + (b^+ - b)\eta|-1\rangle\langle -1| - (1|(b^+ - b))}$, the radio-frequency part becomes

$$\begin{aligned} UH_U = & \Omega_g(|-1\rangle\langle 0'|e^{-\eta(b^+ - b)}e^{i\delta t} \\ & + e^{i\theta}|1\rangle\langle 0'|e^{\eta(b^+ - b)}e^{i\delta t} + \text{h.c.}) \end{aligned}$$

To first order in the Lamb-Dicke parameter, for the D qubit we obtain the Hamiltonian

$$\sqrt{2}\eta\Omega_g(|D\rangle\langle 0'|e^{i\delta t} - \text{h.c.})(b^+ - b)$$

To zeroth order, we obtain a coupling between $|B\rangle$ and $|0'\rangle$; however, this term is ignored because the coupling, Ω_g , is much smaller than the energy gap, which is of order Ω . For the B qubit, we get the analogous Hamiltonian (Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 March; accepted 16 June 2011.

- Blatt, R. & Wineland, D. Entangled states of trapped atomic ions. *Nature* **453**, 1008–1015 (2008).
- Friedenauer, A. Schmitz, H. Glueckert, J. T., Porras, D. & Schaetz, T. Simulating a quantum magnet with trapped ions. *Nature Phys.* **4**, 757–761 (2008).
- Kim, K. *et al.* Quantum simulation of frustrated Ising spins with trapped ions. *Nature* **465**, 590–593 (2010).
- Gerritsma, R. *et al.* Quantum simulation of the Dirac equation. *Nature* **463**, 68–71 (2010).
- Johanning, M. Varón, A. F. & Wunderlich, C. Quantum simulations with cold trapped ions. *J. Phys. B* **42**, 154009 (2009).
- Minter, F. & Wunderlich, C. Ion-trap quantum logic using long-wavelength radiation. *Phys. Rev. Lett.* **87**, 257904 (2001).
- Ospelkaus, C. *et al.* Trapped-ion quantum logic gates based on oscillating magnetic fields. *Phys. Rev. Lett.* **101**, 090502 (2008).
- Ozeri, R. *et al.* Errors in trapped-ion quantum gates due to spontaneous photon scattering. *Phys. Rev. A* **75**, 042329 (2007).
- Plenio, M. B. & Knight, P. L. Decoherence limits to quantum computation using trapped ions. *Proc. R. Soc. Lond. A* **453**, 2017–2041 (1997).
- Sørensen, A. & Mølmer, K. Entanglement and quantum computation with ions in thermal motion. *Phys. Rev. A* **62**, 022311 (2000).
- Milburn, G. J. Schneider, S. & James, D. F. V. Ion trap quantum computing with warm ions. *Fortschr. Phys.* **48**, 801–810 (2000).
- Wunderlich, C. in *Laser Physics at the Limit* (eds Meschede, D., Zimmermann, C. & Figger, H.) 261–271 (Springer, 2002).
- Wunderlich, C. & Balzer, C. Quantum measurements and new concepts for experiments with trapped ions. *Adv. At. Mol. Phys.* **49**, 293–372 (2003).
- McHugh, D. & Twamley, J. Quantum computer using a trapped-ion spin molecule and microwave radiation. *Phys. Rev. A* **71**, 012315 (2005).
- Wang, S. X. Labaziewicz, J. Ge, Y., Shewmon, R. & Chuang, I. L. Individual addressing of ions using magnetic field gradients in a surface-electrode ion trap. *Appl. Phys. Lett.* **94**, 094103 (2009).
- Johanning, M. *et al.* Individual addressing of trapped ions and coupling of motional and spin states using RF radiation. *Phys. Rev. Lett.* **102**, 073004 (2009).
- Häffner, H. *et al.* Robust entanglement. *Appl. Phys. B* **81**, 151–153 (2005).
- Kielinski, D. *et al.* A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
- Home, J. P. *et al.* Complete methods set for scalable ion trap quantum information processing. *Science* **325**, 1227–1230 (2009).
- Viola, L. & Lloyd, S. Dynamical suppression of decoherence in two-state quantum systems. *Phys. Rev. A* **58**, 2733–2744 (1998).
- Rabl, P. *et al.* Strong magnetic coupling between an electronic spin qubit and a mechanical resonator. *Phys. Rev. B* **79**, 041302 (2009).
- Biercuk, M. J. *et al.* Optimized dynamical decoupling in a model quantum memory. *Nature* **458**, 996–1000 (2009).
- Bluhm, H. *et al.* Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200 μs . *Nature Phys.* **7**, 109–113 (2011).
- Vitanov, N. V. Fleischhauer, M. Shore, B. W. & Bergmann, K. Coherent manipulation of atoms and molecules by sequential laser pulses. *Adv. At. Mol. Phys.* **46**, 55–190 (2001).
- Sørensen, J. *et al.* Efficient coherent internal state transfer in trapped ions using stimulated Raman adiabatic passage. *New J. Phys.* **8**, 261 1–11 (2006).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Specht, H. P. *et al.* A single-atom quantum memory. *Nature* **473**, 190–192 (2011).
- Simon, C. *et al.* Quantum memories. *Eur. Phys. J. D* **58**, 1–22 (2010).
- Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
- Hannemann, T. *et al.* Self-learning estimation of quantum states. *Phys. Rev. A* **65**, 050303 1–4 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Technical help with the microwave set-up by T. F. Gloger is acknowledged. We acknowledge support by the Bundesministerium für Bildung und Forschung (FK 01BQ1012 and P3352014), the Deutsche Forschungsgemeinschaft, the European Commission under the STREP PICC, the German-Israeli Foundation, secunet AG and the Alexander von Humboldt Foundation.

Author Contributions N.T., I.B., M.J., A.F.V. and Ch.W. contributed to the experiment, the analysis of experimental and theoretical results, and the writing of the manuscript. A.R. and M.B.P. contributed to the theory, the analysis of theoretical and experimental results, and the writing of the manuscript. A.R. and Ch.W. had the idea for theory and experiment.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Ch.W. (wunderlich@physik.uni-siegen.de) or A.R. (alex.retzker@uni-ulm.de).

METHODS

The experimental system. The atomic hyperfine states of the electronic ground state of the $^{171}\text{Yb}^+$ ions used in this work can be mapped to the general scheme shown in Fig. 1 as follows: $|F=0\rangle \leftrightarrow |0\rangle$, $|F=1, m_F=-1\rangle \leftrightarrow |-1\rangle$, $|F=1, m_F=+1\rangle \leftrightarrow |+1\rangle$, $|F=1, m_F=0\rangle \leftrightarrow |0'\rangle$ (ref. 30).

For STIRAP, two microwave fields with Gaussian amplitude envelopes shifted in time relative to each other are used to drive the $|-1\rangle \leftrightarrow |0\rangle$ and $|+1\rangle \leftrightarrow |0\rangle$ transitions, respectively (Fig. 1).

The effectiveness of the STIRAP process for creating dressed states and transferring them into the final state has been investigated as a function of the parameters characterizing the pulse sequence (see ref. 25 for optical STIRAP). The sequence is divided into discrete time increments $\Delta t = 1/f_{\Omega}N_i$, where N_i is a positive integer, $f_{\Omega} = \Omega_p/2\pi$ and Ω_p is the peak Rabi frequency. The width of the Gaussian pulses is given by N/f_{Ω} , where N is an integer. The STIRAP sequence is robust to variations of these parameters: we investigated the range $10 \leq N_i \leq 40$ for $N = 10$ and found no variation in the effectiveness of the pulse sequence; that is, the overall fidelity of initial preparation of $|0\rangle$, preparation of $|D\rangle$ and read out stays constant (within statistical variations) at a value of about 93%.

When measuring sequences with pulse widths varying in the range $2 \leq N \leq 20$, equally good results were obtained for $N \geq 4$. The separation in time, s_i/f_{Ω} , of the two Gaussian pulses was varied over the range $0 \leq s_i/f_{\Omega} \leq 40/f_{\Omega}$, for a pulse width with $N = 10$ to obtain a plateau of high effectiveness ($\sim 93\%$) for values of s_i in the interval $10 \leq s_i \leq 20$. The best performance is obtained for equal detunings $\Delta_+ = \Delta_-$, ideally $\Delta_+ = 0 = \Delta_-$, and for small relative detunings. For $|\delta| = |\Delta_+ - \Delta_-| < 0.1\Omega$, an experimental investigation yielded no statistically significant variation of the dressed-state preparation fidelity.

To prepare, for example, the state $|D\rangle$, the population is first transferred from the initial state $|0\rangle$ to the atomic state $|-1\rangle$ by applying a microwave π -pulse. The first half of the STIRAP sequence then transfers the atomic population to the dressed state $|D\rangle$.

A holding period in the evolution of the two STIRAP microwave fields is introduced at the crossing point of the amplitude envelopes of the two pulses. At that time, the occupation of $|D\rangle$ is at its maximum. Any dephasing of $|D\rangle$ or transitions to other states during the holding time, T , gives rise to imperfect population transfer during the second half of the STIRAP sequence that transfers the system to the atomic state $|+1\rangle$.

Probing the state after the complete STIRAP sequence does not distinguish between $|-1\rangle$ and $|+1\rangle$, as both yield bright results on final detection of resonance fluorescence. Therefore, a π -pulse swaps the populations of $|+1\rangle$ and $|0\rangle$ before final detection. Thus, a dark result indicates a successful STIRAP transfer between atomic states and dressed states, and a lifetime T of the dressed state prepared during the holding period (Fig. 2).

To record Rabi oscillations between dressed states, we first prepare state $|B\rangle$. Then a radio-frequency pulse in resonance with $|0'\rangle \leftrightarrow |\pm 1\rangle$ (Fig. 1a) is applied during the holding time, T , in the evolution of STIRAP. This radio-frequency pulse induces Rabi oscillations between $|B\rangle$ and $|0'\rangle$. After the Rabi pulse, the STIRAP pulse sequence is completed and the population of the atomic state $|0\rangle$ is probed as described above. If during the radio-frequency Rabi pulse state $|0'\rangle$ is populated, then the second part of the STIRAP sequence has no effect; that is, $|0\rangle$ is not populated at the end of the experimental sequence.

Single-qubit gate. We refer to $\{|0'\rangle, |D\rangle\}$ as the D qubit and to $\{|0'\rangle, |B\rangle\}$ as the B qubit in what follows. The Hamiltonian reads:

$$\begin{aligned} H_{\text{sqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_-t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_+t} + \text{h.c.}) \\ & + \Omega_g \cos(\lambda_0 t)(|-1\rangle\langle 0'| + \text{h.c.}) \\ & + \Omega_g \cos(\lambda_0 t + \phi)(|1\rangle\langle 0'| + \text{h.c.}) \end{aligned}$$

Here, ω_0 is the zero-field hyperfine splitting, $\omega_{\pm 1}$ are the microwave frequencies with their initial phase difference θ , λ_0 is the radio frequency, and ϕ the initial phase difference between the radio frequency fields. For $\theta = 0$ and $\phi = \pi$ we obtain the D qubit, and for $\theta = \pi$ and $\phi = 0$ we obtain the B qubit. In the rotating-wave approximation we get

$$\begin{aligned} H_{\text{sqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_-t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_+t} + \text{h.c.}) \\ & + \Omega_g(|-1\rangle\langle 0'|e^{-i\lambda_0 t} - e^{i\theta}|1\rangle\langle 0'|e^{i\lambda_0 t} + \text{h.c.}) \end{aligned}$$

with $\theta = 0$ for the D qubit and $\theta = \pi$ for the B qubit. If we set $\omega_{-1} = \omega_0 + \lambda_0$ and $\omega_{+1} = \omega_0 - \lambda_0$, then by moving to the interaction picture with respect to the time-independent part we find, for the D qubit, that

$$\begin{aligned} H_{\text{sqg}}^I = & \sqrt{2}\Omega(|B\rangle\langle 0| + \text{h.c.}) + \sqrt{2}\Omega_g(|D\rangle\langle 0'| + \text{h.c.}) \\ = & \frac{\Omega}{\sqrt{2}}|u\rangle\langle u| - \frac{\Omega}{\sqrt{2}}|d\rangle\langle d| + \sqrt{2}\Omega_g(|D\rangle\langle 0'| + \text{h.c.}) \end{aligned}$$

The first two terms shift states $|u\rangle$ and $|d\rangle$ (which both contain $|B\rangle$) away from $|D\rangle$, and the second part creates the gate. The interactions created by the two microwave fields with Rabi frequency Ω and the radio-frequency fields of frequency Ω_g (Fig. 1a) permit the implementation of this single-qubit gate. The two radio-frequency fields have a phase difference of π at the initial time of the pulse, which realizes the coupling to $|D\rangle$ and not to $|B\rangle$.

Alternatively, we may work with the B qubit by choosing the same initial phase for the two radio-frequency fields; that is, one radio-frequency field is sufficient to couple to $|B\rangle$. We find that

$$\begin{aligned} H_{\text{sqg}}^I = & \sqrt{2}\Omega(|D\rangle\langle 0| + \text{h.c.}) + \sqrt{2}\Omega_g(|B\rangle\langle 0'| + \text{h.c.}) \\ = & \frac{\Omega}{\sqrt{2}}|u\rangle\langle u| - \frac{\Omega}{\sqrt{2}}|d\rangle\langle d| + \sqrt{2}\Omega_g(|B\rangle\langle 0'| + \text{h.c.}) \end{aligned}$$

Multiqubit gate. The levels and fields of Fig. 1b (for the D qubit) are described by the Hamiltonian

$$\begin{aligned} H_{\text{mqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_-t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_+t} + \text{h.c.}) \\ & + \Omega_g \cos[(\lambda_0 - \delta)t](|-1\rangle\langle 0'| + \text{h.c.}) \\ & + \Omega_g \cos[(\lambda_0 - \delta)t + \theta + \pi](|1\rangle\langle 0'| + \text{h.c.}) \\ & + vb^+b|1\rangle\langle 1| + \lambda(|-1\rangle\langle -1| - |1\rangle\langle 1|)(b + b^+) \end{aligned}$$

where λ is proportional to the magnetic gradient as in ref. 6, δ is the detuning of the radio frequency field, $\theta = 0$ for the D qubit and $\theta = \pi$ for the B qubit. In the rotating wave approximation, we get:

$$\begin{aligned} H_{\text{mqg}} = & \omega_0|0\rangle\langle 0| + \lambda_0(|1\rangle\langle 1| - |-1\rangle\langle -1|) \\ & + \Omega(|-1\rangle\langle 0|e^{-i\omega_-t} + e^{i\theta}|1\rangle\langle 0|e^{-i\omega_+t} + \text{h.c.}) \\ & + \Omega_g(|-1\rangle\langle 0'|e^{-i(\lambda_0 - \delta)t} + e^{i\theta}|1\rangle\langle 0'|e^{-i(\lambda_0 - \delta)t} + \text{h.c.}) \\ & + vb^+b|1\rangle\langle 1| + \lambda(|-1\rangle\langle -1| - |1\rangle\langle 1|)(b + b^+) \end{aligned}$$

In the interaction picture with respect to the microwaves, and after the polaron transformation $U = e^{\eta|+1\rangle\langle +1|(b^+ - b)}e^{-\eta|-1\rangle\langle -1|(b^+ - b)}$, the radio-frequency part becomes

$$\begin{aligned} UH_{\text{mqg}}U^+ = & \Omega_g(|-1\rangle\langle 0'|e^{-\eta(b^+ - b)}e^{i\delta t} \\ & + e^{i\theta}|1\rangle\langle 0'|e^{\eta(b^+ - b)}e^{i\delta t} + \text{h.c.}) \end{aligned}$$

To first order in the Lamb-Dicke parameter, for the D qubit we obtain the Hamiltonian

$$\sqrt{2}\eta\Omega_g(|D\rangle\langle 0'|e^{i\delta t} - \text{h.c.})(b^+ - b)$$

To zeroth order, we obtain a coupling between $|B\rangle$ and $|0'\rangle$; however, this term is ignored because the coupling, Ω_g , is much smaller than the energy gap, which is of order Ω . For the B qubit, we get the analogous Hamiltonian

$$\sqrt{2}\eta\Omega_g(|B\rangle\langle 0'|e^{i\delta t} - \text{h.c.})(b^+ - b)$$

When using a microwave frequency comb to dress many non-degenerate qubits, small errors due to higher-order effects may be introduced and the implementation of gates may lead to off-resonance transitions and, hence, phase shifts in other ions. However, it is important to note that the size of this error is exactly the same as for magnetic-gradient-induced coupling⁶. This is due to the fact that the microwave fields almost completely decouple the $\{|u\rangle, |d\rangle\}$ subspace from the qubit subspace ($\{|D\rangle, |0'\rangle\}$) as long as Ω is much larger (typically an order of magnitude) than the radio-frequency coupling, Ω_g .

Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection

Ioan Mihai Miron¹, Kevin Garello¹, Gilles Gaudin², Pierre-Jean Zermatten², Marius V. Costache¹, Stéphane Auffret², Sébastien Bandiera², Bernard Rodmacq², Alain Schuhl² & Pietro Gambardella^{1,3,4}

Modern computing technology is based on writing, storing and retrieving information encoded as magnetic bits. Although the giant magnetoresistance effect has improved the electrical read out of memory elements, magnetic writing remains the object of major research efforts¹. Despite several reports of methods to reverse the polarity of nanosized magnets by means of local electric fields^{2,3} and currents^{4–6}, the simple reversal of a high-coercivity, single-layer ferromagnet remains a challenge. Materials with large coercivity and perpendicular magnetic anisotropy represent the mainstay of data storage media, owing to their ability to retain a stable magnetization state over long periods of time and their amenability to miniaturization⁷. However, the same anisotropy properties that make a material attractive for storage also make it hard to write to⁸. Here we demonstrate switching of a perpendicularly magnetized cobalt dot driven by in-plane current injection at room temperature. Our device is composed of a thin cobalt layer with strong perpendicular anisotropy and Rashba interaction induced by asymmetric platinum and AlO_x interface layers^{9,10}. The effective switching field is orthogonal to the direction of the magnetization and to the Rashba field. The symmetry of the switching field is consistent with the spin accumulation induced by the Rashba interaction and the spin-dependent mobility observed in non-magnetic semiconductors^{11,12}, as well as with the torque induced by the spin Hall effect in the platinum layer^{13,14}. Our measurements indicate that the switching efficiency increases with the magnetic anisotropy of the cobalt layer and the oxidation of the aluminium layer, which is uppermost, suggesting that the Rashba interaction has a key role in the reversal mechanism. To prove the potential of in-plane current switching for spintronic applications, we construct a reprogrammable magnetic switch that can be integrated into non-volatile memory and logic architectures. This device is simple, scalable and compatible with present-day magnetic recording technology.

The coupling of spin and orbital angular momenta underlies the magnetic anisotropy properties of ferromagnets. Strong anisotropy allows for permanent, stable storage but also requires stronger magnetic fields to write information to magnetic media. An ideal solution to this problem requires the spin–orbit interaction to be beneficial for both storage and writing purposes. Experiments on non-magnetic semiconductors have revived interest in effective magnetic fields originating in spin–orbit coupling^{11,15–20}. Such fields relate the spin of an electron to its momentum, converting a charge current into a source of spin polarization even in the absence of magnetism. More recently, spin–orbit fields have been predicted^{21,22} and observed in ferromagnets lacking structural inversion symmetry^{6,9}, where the *s*–*d* exchange interaction between charge carriers and localized *d* electrons mediates the effect of the spin–orbit fields on the magnetization. The possibility of generating strong spin–orbit fields in ferromagnetic metals (FMMs) is particularly interesting for applications, owing to the robust Curie temperature and perpendicular magnetic anisotropy (PMA) afforded

by such systems. So far, however, the centrosymmetric lattice properties of FMMs restrict this possibility to generation of the Rashba effect²³, as inversion asymmetry can only be achieved by sandwiching an FMM layer between two dissimilar interfaces, one of which is usually a heavy, non-magnetic element^{9,24}. Owing to the Rashba interaction, electrons flowing in the plane of a conductor with asymmetric interfaces experience an effective magnetic field $\mathbf{B}_R = \alpha_R(\hat{\mathbf{z}} \times \mathbf{j})$, where α_R is the Rashba constant, $\hat{\mathbf{z}}$ is a unit vector along the perpendicular axis (the *z* direction) and \mathbf{j} is the vector current density. As the interface effects responsible for the Rashba interaction also induce PMA, \mathbf{B}_R is orthogonal to the easy magnetization axis. Therefore, the Rashba field by itself cannot be used to control magnetization reversal in FMMs⁹.

Here we demonstrate up–down magnetic switching of a cobalt dot characterized by strong PMA and Rashba interaction induced by an in-plane current. Bipolar reversal of the magnetization is achieved by injecting in-plane current pulses of either positive or negative sign parallel to a static magnetic field of moderate amplitude. This effect is discussed in terms of the symmetry of the current-induced spin–orbit fields acting on the cobalt layer. Switching occurs in single-layer FMMs independently of the domain configuration of the sample and without the need for a polarizer layer. We show that this phenomenon leads to innovative device architectures for spintronics.

Each of our samples consists of a 500 nm × 500 nm cobalt dot sandwiched between a 3-nm-thick platinum layer (below) and a 1.6-nm-thick AlO_x cap (above; Fig. 1a). The thickness of the cobalt layer is chosen to be 0.6 nm to maximize the Rashba effect due to structure inversion asymmetry at the platinum and AlO_x interfaces⁹. The AlO_x and cobalt layers are each patterned into a square dot, whereas the platinum layer is etched into a cross-shaped structure for current injection and Hall voltage measurements. The perpendicular magnetization of the dot, M_z , is measured in terms of the anomalous Hall resistance, which varies by $R_{\text{Hall}} = 0.8 \Omega$ for opposite orientations of M_z . A rotating sample stage is used to apply an external magnetic field, \mathbf{B} , at polar and azimuthal angles θ and ϕ (Fig. 1a). To study the effects of an electric current on M_z , we apply 15-ns-long pulses of variable amplitude supplied by a voltage generator, using the circuit shown in Fig. 1b. All measurements are performed at room temperature.

The central result of this work is shown in Fig. 1c, d, where we compare field-induced and current-induced reversal of M_z . We first perform a standard magnetization measurement in the absence of current pulses, recording M_z during a single sweep of \mathbf{B} applied parallel to the current injection line (Fig. 1c, solid line). The field is intentionally tilted off-plane by 2° ($\theta = 92^\circ$, $\phi = 0^\circ$) to prevent the formation of magnetic domains. Magnetization reversal occurs at the coercive field $|B_c| = 300$ mT, owing to the residual component, B_z , parallel to the out-of-plane easy axis of the dot. We then repeat this magnetic field sweep, but stop at each value of B and inject a positive (set) current pulse and negative (reset) current pulse, measuring M_z after each pulse. The black and red points (Fig. 1d) indicate the orientation of M_z after the alternating injection of set and reset pulses of amplitude

¹Catalan Institute of Nanotechnology (ICN-CIN2), E-08193 Barcelona, Spain. ²SPINTEC, UMR-8191, CEA/CNRS/UJF/GINP, INAC, F-38054 Grenoble, France. ³Departament de Física, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain. ⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), E-08010 Barcelona, Spain.

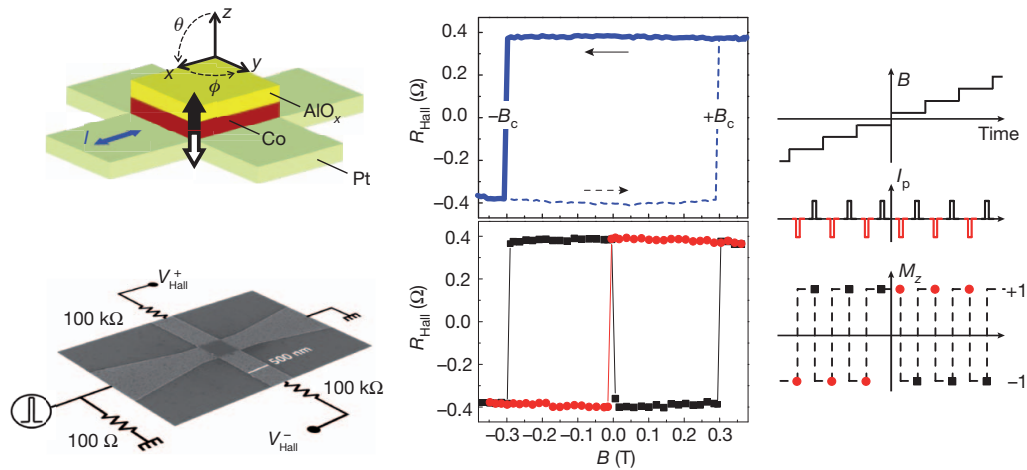


Figure 1 | Device schematic and current-induced switching. **a**, Hall cross geometry. Black and white arrows indicate the ‘up’ and ‘down’ equilibrium magnetization states of the cobalt layer, respectively. **b**, Scanning electron micrograph of the sample and electric circuitry used in the measurements. V_{Hall}^+ and V_{Hall}^- represent the two terminals for the Hall voltage measurements. **c**, M_z measured by the anomalous Hall resistance as a function of applied field, B . **d**, M_z

$I_p = 2.6$ mA. This plot shows two striking effects, illustrated in Fig. 1e. First, black and red points that represent magnetizations of opposite sign indicate switching of M_z from up to down and vice versa following every pulse of current. Second, the sense of switching reverses as B goes through zero. Positive and negative sweeps of B give identical results. We note that pulse-induced magnetization reversal occurs in the entire bistability range of the FMM delimited by B_c , down to $B \approx 5$ mT. This behaviour is independent of the domain configuration of the sample (Supplementary Information) and is remarkably different from that expected for known magnetic interactions. For example, the in-plane Rashba and Oersted fields, as well as Joule

measured after the injection of positive (black squares) and negative (red circles) current pulses of amplitude $I_p = 2.58$ mA. The data are reported during a single sweep of B , corresponding to the solid line in **c**. **e**, Schematic of the pulse sequence and magnetization measurements. In both **c** and **d**, B is applied at $\theta = 92^\circ$, parallel to the current direction ($\phi = 0^\circ$). The 2° offset with respect to the ideal in-plane direction is used to define the residual component B_z unambiguously.

heating, would tend to favour a demagnetized state that is not compatible with the observed deterministic switching⁹, whereas precessional switching does not depend on the current direction²⁵.

Measurements performed over a range of currents offer further insight into the switching phenomenon (Fig. 2). Although weak current pulses have no effect, starting at around $I_p = 1.3$ mA (Fig. 2a) we observe a gradual reduction in the coercive field (Fig. 2c, red triangles). This behaviour can be understood by noting that the combination of a weak current-induced effect and the residual component B_z assists magnetization reversal towards the equilibrium direction, parallel to B_z . At $I_p \approx 1.9$ mA, however, we observe the onset of current-induced

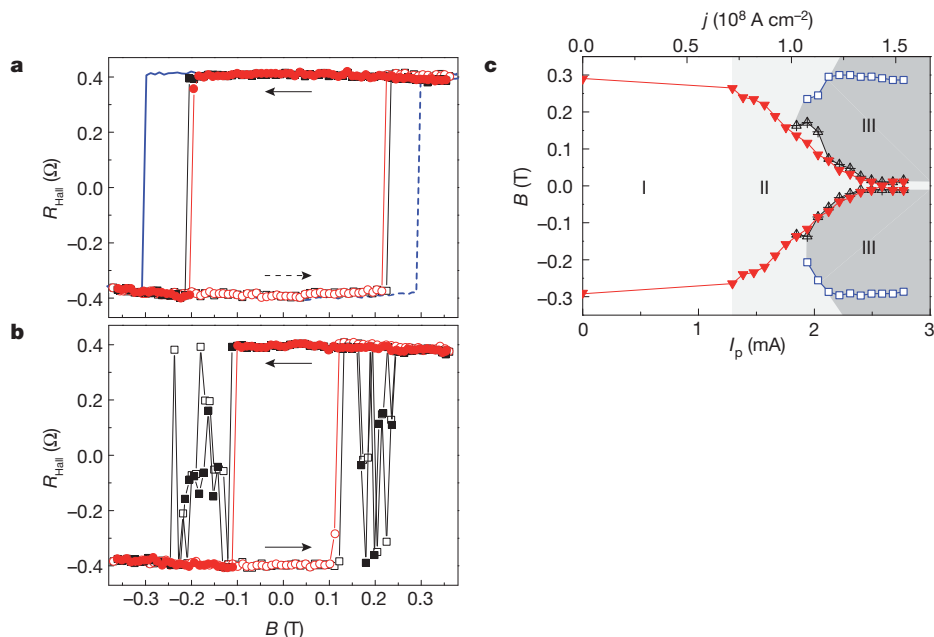


Figure 2 | Switching efficiency as a function of current amplitude. **a**, **b**, M_z measured after injection of positive (black squares) and negative (red circles) current pulses of amplitude $I_p = 1.57$ mA (**a**) and $I_p = 1.94$ mA (**b**). Filled symbols indicate data recorded during a $+B \rightarrow -B$ sweep and open symbols indicate data recorded during a $-B \rightarrow +B$ sweep, as shown by the arrows. The solid and dashed lines in **a** represent M_z as a function of field. **c**, Switching efficiency as a function of pulse amplitude and applied magnetic field. Region I:

conventional field-induced magnetization reversal occurs at $B = B_c$. Region II: assisted reversal. Red triangles indicate the minimum external field required to reverse M_z parallel to B_z . Region III: pulse-induced switching. Black triangles indicate the minimum field at which positive current pulses reverse the magnetization antiparallel to B_z . The maximum field at which switching is observed (blue open squares) coincides with the coercivity of the dot.

switching events (Fig. 2b), which become systematic as I_p is further increased above 2.3 mA (Fig. 1d). The high-field switching limit (Fig. 2c, blue squares) gradually extends up to $B = B_c$, indicating that the torque working against the action of B_z becomes stronger as I_p increases. At the same time, the minimum in-plane field component necessary to trigger the reversal decreases to about 5 mT (Fig. 2c, region III). Furthermore, we find that the critical switching current scales with the lateral dimensions of the cobalt dot and the inverse pulse duration (Supplementary Information).

A systematic study of current-induced magnetization reversal as a function of applied field direction (Fig. 3) shows that the switching torque is maximal when \mathbf{B} is applied parallel to \mathbf{j} and decreases away from this orientation following a $\sin(\phi - \pi/2)$ function. Combining the above information, we observe that the action of the current on the magnetization is equivalent to that of an effective perpendicular magnetic field $\mathbf{B}_{Sz} \approx \mathbf{B}_R \times \mathbf{B} \approx (\hat{\mathbf{z}} \times \mathbf{j}) \times \mathbf{B}$, which changes sign on reversal of either \mathbf{B} or \mathbf{j} (Fig. 3c). In an FMM, however, relative to the internal exchange field produced by the local magnetization, the external field, \mathbf{B} , has no direct influence on the spin polarization of the conduction electrons. Therefore, \mathbf{B}_{Sz} shall be considered the perpendicular component of an effective magnetic field proportional to $\mathbf{B}_R \times \mathbf{M}$, where the orientation of \mathbf{M} is initially determined by \mathbf{B} .

Although there is no theory predicting the magnetization reversal mechanism reported here, we observe that there are at least two effects that can induce spin accumulation parallel to $\mathbf{B}_R \times \mathbf{M}$. We describe the first in analogy to phenomena observed in non-magnetic semiconductors. Faraday measurements performed on InGaAs heterostructures have shown that a magnetic field applied parallel to \mathbf{j} induces a spin polarization component perpendicular to the plane of the current¹¹. This phenomenon has been explained by the combined action of \mathbf{B}_R and spin-dependent electron mobility, which leads to a net perpendicular spin accumulation proportional to the cross product $\mathbf{B}_R \times \mathbf{B}$ (ref. 12). We emphasize that this term has the same symmetry, current and field dependence as the effective switching field. This

analogy suggests that a similar mechanism might be at work in our case, considering that the spin-dependent conductivity required to explain the effect in InGaAs is a common feature of FMMs²⁶. Following this argument (see Supplementary Information for more details), the accumulation of out-of-plane spin polarization in an FMM requires a non-zero in-plane magnetization component parallel to \mathbf{j} , induced by a small external field. The final effect, mediated by s - d exchange, is equivalent to that of a downward or upward effective field, \mathbf{B}_{Sz} , parallel to the z component of $\mathbf{B}_R \times \mathbf{M}$, in agreement with our measurements.

The second mechanism is based on the absorption of the spin current produced by the spin Hall effect (SHE) in platinum^{13,14} and diffusing into the cobalt layer. The absorption of this current, polarized along the y direction, is equivalent to a torque acting on the cobalt magnetization that has the same symmetry as the effect reported here. Estimates based on the spin Hall angle in platinum, however, show that the SHE is not sufficient to account for the intensity of the switching field (Supplementary Information). Moreover, we find that samples with stronger PMA and a larger degree of aluminium oxidation switch more easily than do magnetically softer dots with equally thick cobalt and platinum layers (Supplementary Figs 8 and 9). This observation is not compatible with a simple explanation based on the SHE. Nevertheless, we cannot exclude either a partial contribution by the SHE to switching or a more complex mechanism involving the SHE. For example, the vertical gradient of the Rashba field at the cobalt-platinum interface might induce a supplementary spin current that adds to the SHE in platinum.

Although this work calls for a detailed theory of non-equilibrium spin-orbit phenomena in FMMs, it is important to distinguish two practical features of our results. First, when sufficient current and field are applied in the same direction, switching is robust and limited only by the intrinsic bistability of the FMM. Because neither the external field value nor its exact orientation is critical, a constant magnetic field produced by a macroscopic permanent magnet can be used to ensure

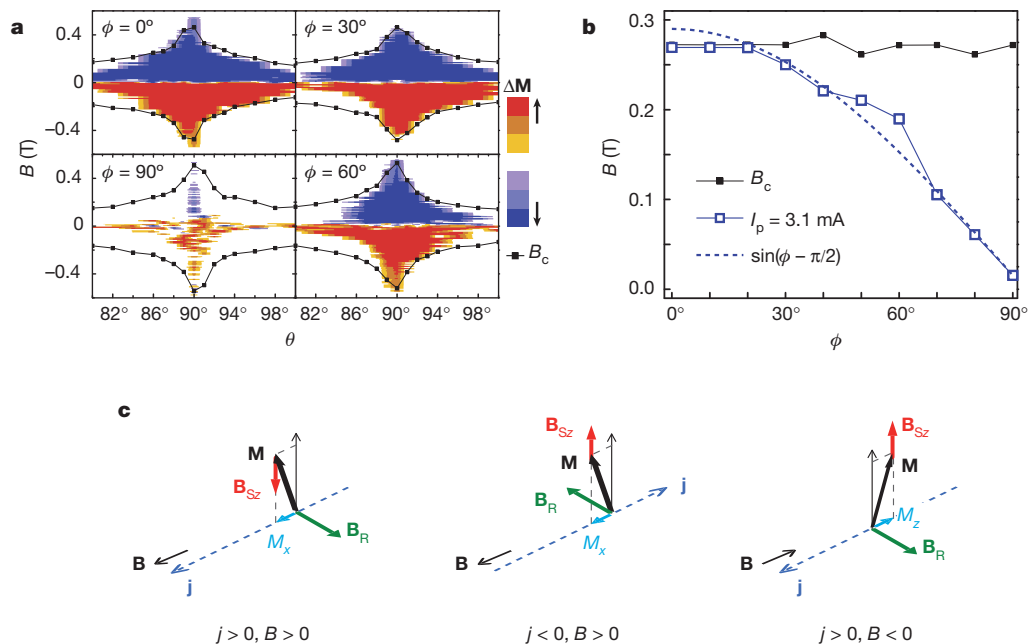


Figure 3 | Dependence of switching on applied field direction. **a**, Diagrams representing M_z reversal regions, showing the intensity and orientation of the applied magnetic field, \mathbf{B} , at constant current ($I_p = 3.3$ mA). Colours indicate the switching sense: red, upwards; blue, downwards. Filled symbols show the coercive field at each angle. When \mathbf{B} is aligned with the current ($\phi = 0^\circ$), the switching areas extend throughout the entire bistability region delimited by B_c , with the exception of a narrow area near zero field. As \mathbf{B} is rotated away from the current, the sizes of the switching areas gradually decrease. When \mathbf{B} is

perpendicular to the current ($\phi = 90^\circ$), switching disappears and is replaced by random nucleation. **b**, Azimuthal dependence of the switching efficiency at constant current ($I_p = 3.1$ mA, $\theta = 87^\circ$). Open symbols represent the largest field against which switching is observed; filled symbols show the coercive field. The dashed line is a sine function. **c**, Directions of the effective switching field, \mathbf{B}_{Sz} , relative to \mathbf{j} , the in-plane magnetization component (M_x) induced by \mathbf{B} , and the Rashba field, \mathbf{B}_R . The magnetization switches from 'up' to 'down' for $j > 0$, $B > 0$ and $j < 0$, $B < 0$, and from 'down' to 'up' for $j > 0$, $B < 0$ and $j < 0$, $B > 0$.

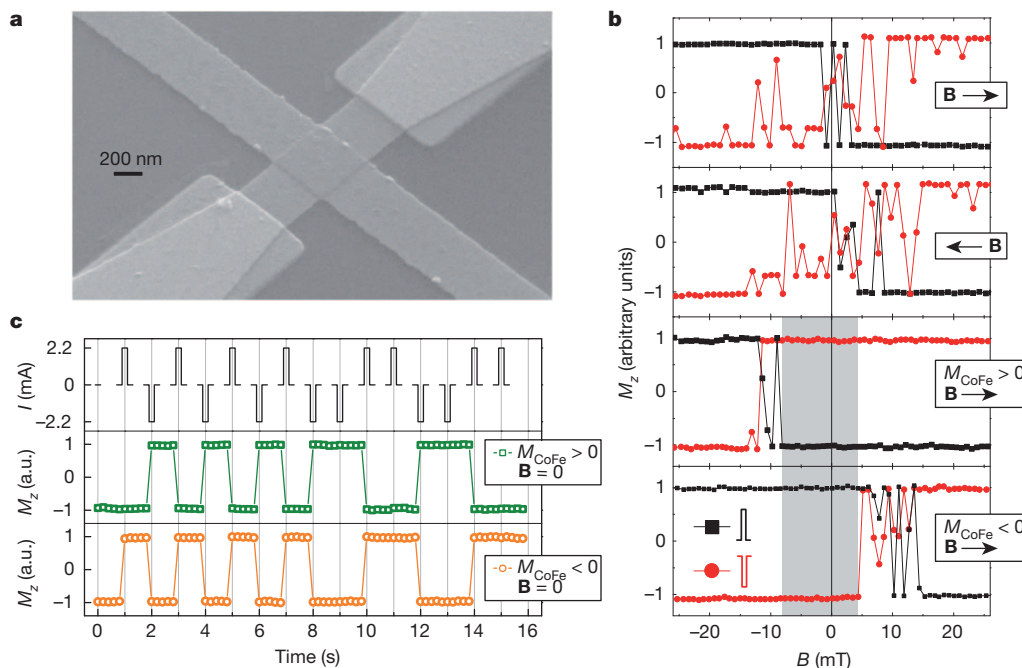


Figure 4 | Prototype of a reconfigurable ferromagnetic switch. **a**, Scanning electron micrograph of a device comprising two 50-nm-thick CoFe bar magnets parallel to the current injection line (bottom left and upper right corners) and an AlO_x-cobalt-platinum dot (centre). **b**, Top two panels: switching curve of a device without CoFe magnets around the zero field region. M_z is measured after injection of positive (black squares) and negative (red circles) pulses. No difference between positive and negative field sweeps is observed apart from random nucleation events near $B = 0$. Bottom two panels:

switching curve of a device including on-chip magnets. The shaded areas show that bipolar switching extends to zero external field. Positive and negative field sweeps reveal the hysteretic behaviour of the CoFe magnets as well as the inversion of the switching polarity near $B = 0$. **c**, Controlled switching sequence of the device in zero field after setting positive (green squares) and negative (orange circles) CoFe magnetization. The switching current is $I_p = 2.5$ mA for pulses 9 ns long. a.u., arbitrary units.

bidirectional current switching of many independent magnetic bits. Second, the area separating the two opposite switching zones near $B = 0$ is very narrow, meaning that very small applied fields suffice to induce reversal. This feature prompts an approach for applications. Small permanent magnets can be fabricated on-chip, close to or on top of each active dot. In this way, not only can the dots be switched, but their response to the current direction can also be inverted by reversing the polarization of the magnets.

The ability to change the operation mode of a device is extremely valuable for programmable magnetic logic applications^{27,28}. This motivated us to fabricate a small prototype switch by depositing two CoFe bars on top of the current injection lines near the cobalt dot (Fig. 4a). The shape anisotropy of the CoFe magnets ensures the alignment between their stray field and current lines. Figure 4b shows the low-field region of a typical switching curve, comparing results for a device without magnets (the two upper panels) with those of a device with CoFe bars (the two bottom panels). In the first case, as B approaches zero, switching is replaced by random nucleation events. In the second case, owing to the additional stray field of the magnets, the switching curve shows hysteretic behaviour extending to the zero-field region (shaded area). We further tested zero-field switching using a complex sequence of current pulses (Fig. 4c). After the magnetization of the CoFe magnets has been saturated in one of two opposite directions ($M_{\text{CoFe}} > 0$ and $M_{\text{CoFe}} < 0$), M_z reverses each time the pulses change polarity. As expected, reversing M_{CoFe} inverts the operation mode of the device.

The ability to switch a single magnetic layer at room temperature using an in-plane current opens the way to a new generation of spintronic devices, combining planar geometry, highly stable perpendicular magnetization, all-electrical write/read out schemes and reconfigurability. The layer structure of our prototype switch is extremely simple, is scalable and is based on materials compatible with present technology. The current density threshold for switching is of the order of

10^8 A cm^{-2} , and is expected to improve with further interface engineering. We note that, because the current is applied in-plane through a thin lateral surface, the absolute current required for switching can be made very small (< 1 mA), providing significant gains in terms of integration and power consumption. Apart from its use in novel device architectures, this type of ferromagnetic switch also has interesting implications for existing technology. For example, a FMM oxide heterostructure could easily be integrated as the storage layer of a magnetic tunnel junction in magnetoresistive random-access memory, to decouple the read and write current paths. In this way, electrical stress of the tunnel barrier during the write process, which is an outstanding problem of the spin transfer approach^{1,28}, could be avoided while maintaining a large tunnel magnetoresistance and read-out sensitivity.

METHODS SUMMARY

The samples were fabricated from aluminium (1.6 nm), cobalt (0.6 nm) and platinum (3 nm) layers deposited on a thermally oxidized silicon wafer by d.c. magnetron sputtering¹⁰. The deposition rates were 0.05 nm s^{-1} (cobalt and aluminium) and 0.1 nm s^{-1} (platinum) at an argon pressure of 2×10^{-3} mbar. After deposition, the samples were oxidized by a 35-s exposure to radio-frequency oxygen plasma at a pressure of 3×10^{-3} mbar and a radio-frequency power of 10 W. This treatment ensures chemical stability of the layers and preserves the strong PMA typical of cobalt-platinum interfaces^{10,29}. The presence of two dissimilar interfaces, of which one is a heavy metal characterized by strong spin-orbit coupling, gives rise to an electric potential gradient across the cobalt layer. This type of structure inversion asymmetry induces a Rashba-type magnetic field parallel to the y direction when the current flows along the x direction^{9,30}. The AlO_x-cobalt-platinum films were patterned by electron beam lithography and ion beam etching into $500 \text{ nm} \times 500 \text{ nm}$ AlO_x-cobalt dots and platinum Hall crosses (Methods). This geometry allows the magnetic state of the dots to be monitored by electric measurements using the anomalous Hall effect. If an electric current is applied along one arm of the cross, the Hall resistance, R_{Hall} , measured along the transverse arm, is proportional to the perpendicular component of the magnetization. Representative magnetization curves of a $500 \text{ nm} \times 500 \text{ nm}$ device are shown in Supplementary Fig. 1. Smaller devices, $200 \text{ nm} \times 200 \text{ nm}$ in size,

were fabricated to probe the scaling of the critical switching current with device size (Supplementary Fig. 5).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 February; accepted 16 June 2011.

Published online 31 July; corrected 11 August 2011 (see full-text HTML version for details).

1. Chappert, C., Fert, A. & Nguyen Van Dau, F. The emergence of spin electronics in data storage. *Nature Mater.* **6**, 813–823 (2007).
2. Ohno, H. *et al.* Electric-field control of ferromagnetism. *Nature* **408**, 944–946 (2000).
3. Chiba, D. *et al.* Magnetization vector manipulation by electric fields. *Nature* **455**, 515–518 (2008).
4. Myers, E. B., Ralph, D. C., Katine, J. A., Louie, R. N. & Buhrman, R. A. Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867–870 (1999).
5. Ralph, D. C. & Stiles, M. D. Spin transfer torques. *J. Magn. Magn. Mater.* **320**, 1190–1216 (2008).
6. Chernyshov, A. *et al.* Evidence for reversible control of magnetization in a ferromagnetic material by means of spin–orbit magnetic field. *Nature Phys.* **5**, 656–659 (2009).
7. Moser, K. *et al.* Magnetic recording: advancing into the future. *J. Phys. D* **35**, R157–R167 (2002).
8. Batra, S., Hannay, J. D., Zhou, H. & Goldberg, J. S. Investigations of perpendicular write head design for 1 Tb/in². *IEEE Trans. Magn.* **40**, 319–325 (2004).
9. Miron, I. M. *et al.* Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. *Nature Mater.* **9**, 230–233 (2010).
10. Rodmacq, B., Manchon, A., Ducruet, C., Auffret, S. & Dieny, B. Influence of thermal annealing on the perpendicular magnetic anisotropy of Pt/Co/AIO_x trilayers. *Phys. Rev. B* **79**, 024423 (2009).
11. Kato, Y. K., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Current-induced spin polarization in strained semiconductors. *Phys. Rev. Lett.* **93**, 176601 (2004).
12. Engel, H.-A., Rashba, E. I. & Halperin, B. I. Out-of-plane spin polarization from in-plane electric and magnetic fields. *Phys. Rev. Lett.* **98**, 036602 (2007).
13. Ando, K. *et al.* Electric manipulation of spin relaxation using the spin Hall effect. *Phys. Rev. Lett.* **101**, 036601 (2008).
14. Liu, L., Moriyama, T., Ralph, D. C. & Buhrman, R. A. Spin-torque ferromagnetic resonance induced by the spin Hall effect. *Phys. Rev. Lett.* **106**, 036601 (2011).
15. Awschalom, D. & Samarth, N. Spintronics without magnetism. *Physics* **2**, 50 (2009).
16. Kato, Y., Myers, R. C., Gossard, A. C. & Awschalom, D. D. Coherent spin manipulation without magnetic fields in strained semiconductors. *Nature* **427**, 50–53 (2004).
17. Silov, A. Yu. *et al.* Current-induced spin polarization at a single heterojunction. *Appl. Phys. Lett.* **85**, 5929–5931 (2004).
18. Ganichev, S. D. *et al.* Electric current-induced spin orientation in quantum well structures. *J. Magn. Magn. Mater.* **300**, 127–131 (2006).
19. Stern, N. P. *et al.* Current-induced polarization and the spin Hall effect at room temperature. *Phys. Rev. Lett.* **97**, 126603 (2006).
20. Meier, L. *et al.* Measurement of Rashba and Dresselhaus spin–orbit magnetic fields. *Nature Phys.* **3**, 650–654 (2007).
21. Manchon, A. & Zhang, S. Theory of nonequilibrium intrinsic spin torque in a single nanomagnet. *Phys. Rev. B* **78**, 212405 (2008).
22. Garate, I. & MacDonald, A. H. Influence of a transport current on magnetic anisotropy in gyrotropic ferromagnets. *Phys. Rev. B* **80**, 134403 (2009).
23. Bychkov, Yu. A. & Rashba, E. I. Properties of a 2D electron gas with lifted spectral degeneracy. *J. Exp. Theor. Phys. Lett.* **39**, 78–81 (1984).
24. Krupin, O. *et al.* Rashba effect at magnetic metal surfaces. *Phys. Rev. B* **71**, 201403(R) (2005).
25. Tudosa, I. *et al.* The ultimate speed of magnetic switching in granular recording media. *Nature* **428**, 831–833 (2004).
26. Campbell, A. & Fert, A. in *Ferromagnetic Materials* Vol. 3 (ed. Wohlfart, E. P.) 747–803 (North Holland, 1982).
27. Ney, A., Pampuch, C., Koch, R. & Ploog, K. H. Programmable computing with a single magnetoresistive element. *Nature* **425**, 485–488 (2003).
28. Dieny, B. *et al.* Spin-transfer effect and its use in spintronic components. *Int. J. Nanotechnol.* **7**, 591–614 (2010).
29. Gambardella, P. *et al.* Giant magnetic anisotropy of single cobalt atoms and nanoparticles. *Science* **300**, 1130–1133 (2003).
30. Pi, U. H. *et al.* Tilting of the spin orientation induced by Rashba effect in ferromagnetic metal layer. *Appl. Phys. Lett.* **97**, 162507 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. O. Valenzuela and S. F. Alvarado for reading the manuscript and for discussions. This work was supported by the European Research Council (StG 203239 NOMAD), the Ministerio de Ciencia y Innovación (ERA-Net EUI2008-03884, MAT2010-15659) and the Agència de Gestió d'Ajuts Universitaris i de Recerca (2009 SGR 695). Samples were patterned at the NANOFAB facility of the Institut Néel (CNRS).

Author Contributions I.M.M., K.G. and P.G. planned the experiment; I.M.M., G.G., P.-J.Z., M.V.C., S.A., S.B. and B.R. fabricated the samples; I.M.M. and K.G. performed the experiments; and I.M.M., K.G. and P.G. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.M.M. (mihai.miron.icn@uab.es) and P.G. (pietro.gambardella.icn@uab.es).

METHODS

The samples were fabricated from aluminium (1.6 nm), cobalt (0.6 nm) and platinum (3 nm) layers deposited on a thermally oxidized silicon wafer by d.c. magnetron sputtering¹⁰. The deposition rates were 0.05 nm s^{-1} (cobalt and aluminium) and 0.1 nm s^{-1} (platinum) at an argon pressure of 2×10^{-3} mbar. After deposition, the samples were oxidized by exposure to radio-frequency oxygen plasma at a pressure of 3×10^{-3} mbar and a radio-frequency power of 10 W for 35 s. This treatment ensures chemical stability of the layers and preserves the strong PMA typical of cobalt–platinum interfaces^{10,29}. The presence of two dissimilar interfaces, of which one is a heavy metal characterized by strong spin–orbit coupling, gives rise to an electric potential gradient across the cobalt layer. This type of structure inversion asymmetry induces a Rashba-type magnetic field parallel to the y direction when the current flows along the x direction^{9,30}.

Patterning. Square magnetic dots were patterned on top of the Hall cross by using two steps of electron beam lithography followed by ion beam milling and wet etching. During the first step, the current injection line is drawn onto the PMMA resist covering the AlO_x –cobalt–platinum layer. After development, 15 nm of titanium are deposited by electron beam evaporation. The lift-off procedure finally defines a 15-nm-thick titanium line. A second step of electron beam lithography (similar to the first one) is used to draw the transverse Hall branches perpendicular to the current injection line. Subsequently, we repeat the resist development, titanium evaporation and lift-off procedures. The device is now composed of two perpendicular, 15-nm-thick titanium lines that intersect. At their intersection,

the total titanium thickness is 30 nm, the sum of the two depositions. Ion beam etching controlled by secondary ion mass spectroscopy is used to etch away the equivalent of 15 nm of titanium. At this stage, the AlO_x –cobalt–platinum layer is trimmed into a cross as the titanium mask has been removed everywhere except at the intersection of the two lines, where 15 nm of Ti remain unetched. We then used a chemically selective wet etching step to remove the oxide and the cobalt layers that are not covered by titanium. At the end of this process, the sample is composed of a platinum cross with a magnetic dot on top. The success of the fabrication technique is monitored by measurements of the anomalous Hall effect: after the chemical etching, the Hall resistance is reduced to approximately 70% of the initial value. The total device resistance was typically 3–5 k Ω . However, we note that most of the resistance drop occurs in the platinum current lines, which are 3 nm thick over a length of approximately 500 μm .

Electrical measurements. The current pulses were generated by a 100-V/2-A pulse generator (Agilent 8114A/001). The rise and fall times of the pulses were approximately 6 ns, as measured by a fast oscilloscope (Agilent DSO81304A) connected in series with the sample. The length of the current pulses was varied between 10 and 100 ns. We note that the pulse length was always longer than the characteristic time-scale for magnetization dynamics (<1 ns) and that its effect is therefore analogous to that of a continuous current. To preserve the shape of the voltage pulse transmitted along the resistive platinum leads, we connected a 100- Ω resistance in parallel with the sample. Two 100-k Ω resistors were placed between the Hall branches and the data acquisition card (Fig. 1) to prevent current flow through the Hall contacts.

Reduced methane growth rate explained by decreased Northern Hemisphere microbial sources

Fuu Ming Kai^{1†}, Stanley C. Tyler^{1†}, James T. Randerson¹ & Donald R. Blake²

Atmospheric methane (CH_4) increased through much of the twentieth century, but this trend gradually weakened until a stable state was temporarily reached around the turn of the millennium^{1,2}, after which levels increased once more³. The reasons for the slowdown are incompletely understood, with past work identifying changes in fossil fuel, wetland and agricultural sources and hydroxyl (OH) sinks as important causal factors^{1,4–8}. Here we show that the late-twentieth-century changes in the CH_4 growth rates are best explained by reduced microbial sources in the Northern Hemisphere. Our results, based on synchronous time series of atmospheric CH_4 mixing and $^{13}\text{C}/^{12}\text{C}$ ratios and a two-box atmospheric model, indicate that the evolution of the mixing ratio requires no significant change in Southern Hemisphere sources between 1984 and 2005. Observed changes in the interhemispheric difference of ^{13}C effectively exclude reduced fossil fuel emissions as the primary cause of the slowdown. The ^{13}C observations are consistent with long-term reductions in agricultural emissions or another microbial source within the Northern Hemisphere. Approximately half ($51 \pm 18\%$) of the decrease in Northern Hemisphere CH_4 emissions can be explained by reduced emissions from rice agriculture in Asia over the past three decades associated with increases in fertilizer application⁹ and reductions in water use^{10,11}.

Several mechanisms have been proposed to explain the long-term slowdown of atmospheric CH_4 , including decreases in source emissions^{4,5,8,10,12}, changes in sink processes⁷, and a stabilization of CH_4 with relatively constant global emissions¹. Using a three-dimensional atmospheric model, Bousquet *et al.*⁵ provide evidence from a Bayesian inversion that the CH_4 slowdown in the 1990s was caused partly by decreases in fossil emissions in the Northern Hemisphere. Carbon and hydrogen isotope measurements have become an important means of determining the magnitude of sources and sinks of atmospheric CH_4 (ref. 13), including assessments of interannual- to centennial-scale trends in the global budget^{5,14–16}. Here we analysed isotope and mixing ratio measurements from the University of California, Irvine (UCI), National Institute of Water and Atmospheric Research (NIWA) and the University of Washington, Stable Isotope Laboratory (SIL) networks^{2,15–17} (Supplementary Table 1). Our goal was to quantify inter-hemispheric differences (IHDs) of CH_4 , $\delta^{13}\text{C}-\text{CH}_4$ and $\delta\text{D}-\text{CH}_4$, and to assess how they constrain recent changes in the global CH_4 budget.

From the mixing ratio measurements, we could see that CH_4 growth rates in both hemispheres declined during 1979–2005 (Supplementary Fig. 1). The IHD of CH_4 (defined as Northern Hemisphere CH_4 minus Southern Hemisphere CH_4) started at approximately 80 ± 20 parts per billion (p.p.b.) during 1979–1981, and gradually increased to

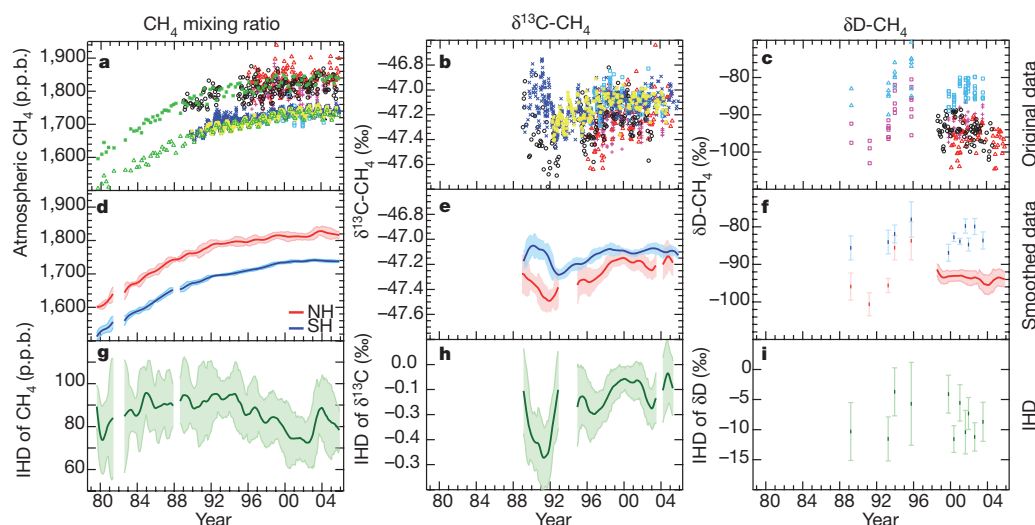


Figure 1 | Long-term trends in atmospheric CH_4 , $\delta^{13}\text{C}-\text{CH}_4$, and $\delta\text{D}-\text{CH}_4$. The left panels show CH_4 ; the middle panels show $\delta^{13}\text{C}-\text{CH}_4$ and the right panels show $\delta\text{D}-\text{CH}_4$. **a–c**, Original measurements used to construct the Northern Hemisphere and Southern Hemisphere time series. The observing stations are Niwot Ridge (black circles), Montana de Oro (red triangles), the Pacific Ocean cruise in the Northern Hemisphere (POCNH; magenta pluses), the UCI network 40°N (green asterisks), the SIL-POCNH (magenta squares), Baring Head (blue crosses), Scott Base (yellow asterisks), the Pacific Ocean cruise in the Southern Hemisphere (POCSH; cyan squares), the UCI network

40°S (green triangles), and the SIL-POCSH (cyan triangles). **d–f**, Twelve-month smoothed trends in the Northern Hemisphere (red line) and Southern Hemisphere (blue line). The smoothed trends were constructed after removing an annual cycle and applying an offset to each station based on hemisphere-specific reference sites. **g–i**, IHDs (green lines) were defined as the difference between the Northern Hemisphere and Southern Hemisphere smoothed time series in **d** to **f**. Error bars are one standard deviation (s.d.), on the basis of all the adjusted observations in a 12-month window centred at each monthly time step. For $\delta\text{D}-\text{CH}_4$, the error bars are one s.d. of the Pacific voyage measurements.

¹Department of Earth System Science, University of California, Irvine, California 92697, USA. ²Department of Chemistry, University of California, Irvine, California 92697, USA. [†]Present addresses: Singapore-MIT Alliance for Research and Technology, S16-05-08, 3 Science Drive 2, 117543 Singapore (F.M.K.); Department of Chemistry, Norco College, Norco, California 92860, USA (S.C.T.).

93 ± 11 p.p.b. during 1991–1995 (Fig. 1). The IHD subsequently decreased from 1995 to 2002, reaching a mean of 79 ± 10 p.p.b. during 2001–2005. For $\delta^{13}\text{C}$, the Southern Hemisphere measurements indicated no significant long-term trend, with the mean during 1989–1993 ($-47.16 \pm 0.09\text{‰}$; mean \pm standard deviation, s.d.) almost the same as the mean during 2001–2005 ($-47.09 \pm 0.01\text{‰}$). In contrast, $\delta^{13}\text{C}$ in the Northern Hemisphere increased from $-47.40 \pm 0.07\text{‰}$ to $-47.19 \pm 0.03\text{‰}$ (Fig. 1e). As a result, the IHD of $\delta^{13}\text{C}$ narrowed from $-0.24 \pm 0.11\text{‰}$ to $-0.10 \pm 0.04\text{‰}$ over this time period (Fig. 1h). Considered separately from other constraints, the trend towards relatively enriched carbon isotope values in the Northern Hemisphere implied (1) a decrease in sources with isotopically depleted signatures (for example, agriculture, landfills or wetlands), (2) an increase in sources that were isotopically enriched (for example, Northern Hemisphere fossil fuel emissions or biomass burning), or (3) an increase in atmospheric removal by reaction with hydroxyl radical (OH) in the Northern Hemisphere.

We also found that $\delta\text{D}-\text{CH}_4$ did not change substantially in the Northern Hemisphere during 1989–2005 (Fig. 1f). Northern Hemisphere δD remained nearly at $-94 \pm 1\text{‰}$ between 1998 and 2005. The IHD of $\delta\text{D}-\text{CH}_4$ ranged from -12.5 to -2.5‰ during the cruises made by the SIL network between 1989 and 1996 (Fig. 1i). Because of the sensitivity of $\delta\text{D}-\text{CH}_4$ to sink processes¹⁵, the relatively constant $\delta\text{D}-\text{CH}_4$ suggested that there was not a substantial long-term change in the OH sink between 1998 and 2005. Other factors potentially influence $\delta\text{D}-\text{CH}_4$, however, and so to estimate quantitatively the impacts of possible changes in the OH sink⁷, we conducted several sensitivity simulations (Supplementary Table 2).

To assess the implications of the measurements described above and to examine recent competitive hypotheses related to the levelling off of CH_4 mixing ratios, we constructed global CH_4 budget scenarios using a two-box atmospheric model¹⁸ (see Supplementary Methods). In a first step, we constructed a control simulation (scenario 1) that included natural and anthropogenic CH_4 sources (Supplementary Fig. 2). In this simulation, anthropogenic emissions associated with fossil fuels and agriculture—two of the largest anthropogenic terms—remained constant after 1980 (Supplementary Fig. 2a and b). We compared model estimates of CH_4 mixing and isotope ratio with observations, which were extended back in time with firn air and ice core measurements of CH_4 (ref. 19) and $\delta^{13}\text{C}-\text{CH}_4$ (ref. 20). The control captured the long-term trends in CH_4 and most of the variations in the CH_4 IHD (Supplementary Fig. 2c and d). Considering CH_4 changes alone, these results appear to suggest that the long-term CH_4 slowdown can be explained by the levelling off of both fossil and agricultural emissions. However, differences between the model estimates and observations of $\delta^{13}\text{C}-\text{CH}_4$ were substantial. For example, the model IHD of $\delta^{13}\text{C}-\text{CH}_4$ remained approximately the same during 1989–2005, whereas the observations showed an increasing trend (a lessening of the IHD (Supplementary Fig. 2f). The differences between the observed and modelled IHD in $\delta^{13}\text{C}-\text{CH}_4$ implied that agricultural and fossil emissions may have had diverging trajectories and provided motivation for exploring the different emissions scenarios described below.

In past work, decreases in fossil fuel emissions have been implicated as an important driver of the methane slowdown^{5,8,12}. To assess this hypothesis using our isotope data, we constructed a second scenario assuming that fossil fuel emissions were solely responsible for the observed CH_4 trends and assuming that all the other sources were the same as in the control. First, we estimated the total flux of CH_4 in the two hemispheres using the smoothed observations (Fig. 1d) and a simple mass-balance inversion during 1984–2005. Figure 2a shows that the Northern Hemisphere flux decreased from 417 ± 10 Tg CH_4 yr⁻¹ during 1984–1985 to 386 ± 5 Tg CH_4 yr⁻¹ during 2004–2005. In parallel, the Southern Hemisphere flux showed no significant change, starting at 148 ± 5 and ending at 150 ± 5 Tg CH_4 yr⁻¹ during this period.

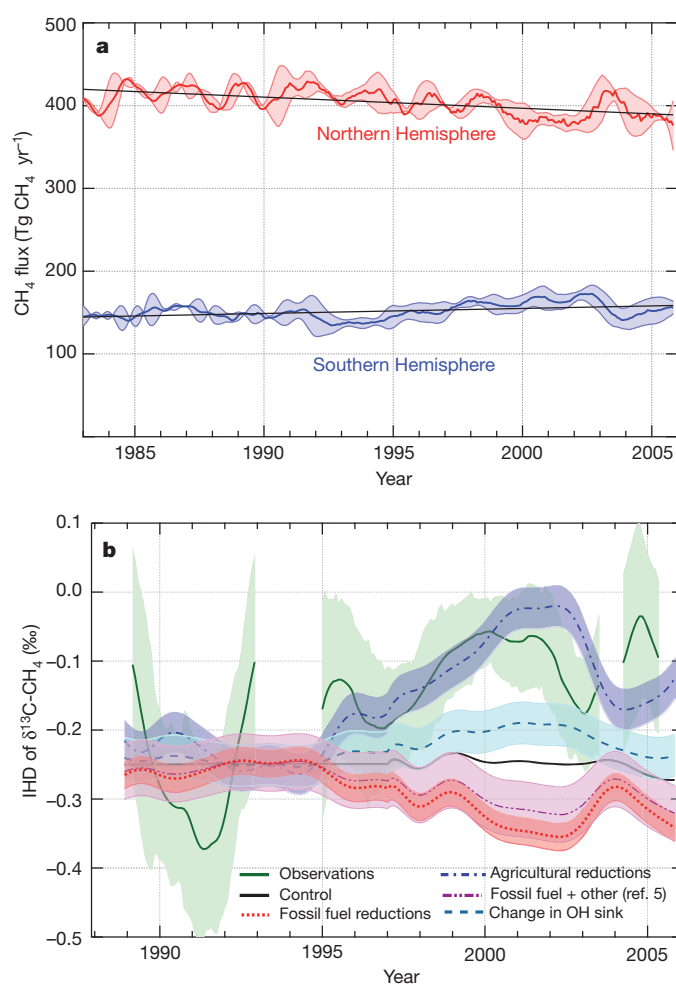


Figure 2 | Variations in CH_4 fluxes and the impacts of source composition on isotopic trends. **a**, The CH_4 fluxes in both hemispheres were obtained from a mass-balance inversion using the mixing ratio observations shown in Fig. 1d. **b**, Comparison of the observed IHD of $\delta^{13}\text{C}$ with model scenarios during 1989–2005. The simulation with decreases in fossil fuel emissions (scenario 2) did not capture the trend in the observed IHD of $\delta^{13}\text{C}$: the $\delta^{13}\text{C}$ IHD in this scenario widened over 1989–2005. In contrast, the simulation with decreases in agricultural emissions (scenario 3) had an increasing trend in the IHD of $\delta^{13}\text{C}$, similar to the observations. A scenario that included decreases in fossil fuels and other sources from ref. 5 (scenario 4) had a negative slope that fell between the control and fossil fuel scenarios. Additional sensitivity simulations with changes in sink processes (scenario 7 as described in the Supplementary Information) show the maximum potential for OH variability to contribute to the observed trend. For the CH_4 fluxes, error bars are one s.d. estimates of the fluxes derived from our inversion in a 12-month window centred at each monthly time step. For comparison, error bars are one s.d. estimates of the observations and model predictions in a 12-month window centred at each monthly time step. The standard deviations of the observations are the same as shown in Fig. 1. For the model predictions, the standard deviations take into account variations in model parameters, including the interhemispheric exchange time, the isotopic fractionation, and the isotopic composition of the sources. The slopes of the regression lines and standard deviation estimates are provided in Supplementary Tables 2 and 4.

For this simulation (scenario 2), we assumed that all of the adjustments to the fluxes in both hemispheres (shown in Fig. 2a) were attributed to changes in fossil fuel emissions—with an isotopic signature specific to this source (Supplementary Table 3). The only difference between this scenario and the control run was the inventory of fossil fuel emissions (Supplementary Figs 2 and 3). As expected, this simulation reproduced most of the CH_4 mixing ratio observations (Supplementary Fig. 3c and d), with decreases in fossil fuel emissions after 1990 (Supplementary Fig. 3a). The decreases in fossil fuel, however, failed to meet

the constraints offered by the measurements of $\delta^{13}\text{C}\text{-CH}_4$ (Supplementary Fig. 3e). Importantly, the $\delta^{13}\text{C}$ IHD in this scenario widened during 1989–2005 (Fig. 2b and Supplementary Fig. 3f). This contrasts with the observed $\delta^{13}\text{C}$ IHD that narrowed during the same time period (Fig. 2b and Supplementary Table 2). In this context, the $\delta^{13}\text{C}$ IHD observations did not support the hypothesis that decreases in fossil fuel emissions were the main contributor to the CH_4 slowdown during 1989–2005.

Agriculture, including animal²¹ and rice⁴ emissions components, is a large term in the contemporary budget⁶ and one that has been subject to considerable modification over the past four decades^{22,23}. To assess the role of agriculture, we created another scenario (scenario 3) in which we assigned the isotope ratio of the Northern Hemisphere and Southern Hemisphere fluxes needed to match the CH_4 observations solely to agricultural sources (Supplementary Fig. 4a and b) following the same approach as described above for the fossil fuel scenario. As expected, the agricultural scenario also matched the mixing ratio observations (Supplementary Fig. 4c and d). In contrast with the fossil fuel scenario, however, the $\delta^{13}\text{C}$ IHD from this run had a positive slope that captured most of the observed trend during 1989–2005 (Fig. 2b and Supplementary Fig. 4f).

We also constructed additional sensitivity and Monte Carlo simulations with our box model examining uncertainties associated with the OH sink, the north–south interhemispheric exchange time for atmospheric mixing, and the isotopic composition of different source terms (as well as potential trends in the isotopic signature of these sources) (Supplementary Tables 2 and 4). Together, these scenarios suggested that our simulation results were robust with respect to our choice of model parameters. Other key uncertainties in our analysis were associated with the representativeness of our reference stations in the Northern and Southern hemispheres, the large variations in CH_4 and $\delta^{13}\text{C}\text{-CH}_4$ observed at the beginning of our time series, and the challenges involved in combining stable isotope observations from several laboratories into one long-term time series (comparisons with other available mixing ratio and isotope observations are provided in the Supplementary Information). The mechanisms causing the large variations in the global methane cycle during the early 1990s have not been fully resolved¹⁶ and may influence the long-term trend in the $\delta^{13}\text{C}\text{-CH}_4$ IHD reported here.

What possible explanations are there for the long-term reductions in microbial sources within the Northern Hemisphere required by the atmospheric isotope observations? Important microbial sources include wetlands, ruminant animals, landfills and rice agriculture. We summarize below evidence that decreasing emissions from rice agriculture probably accounted for some of the observed source reductions in the Northern Hemisphere, using a biogeochemical model to synthesize information from country-level agricultural statistics and recent field studies. Improved management of landfills also may have contributed to Northern Hemisphere source reductions²³. For ruminant animals and wetlands, however, key processes regulating these sources have not changed in a way that is consistent with the long-term decreases in emissions needed to explain the atmospheric observations^{21–25}.

Reductions in rice agriculture emissions have occurred primarily from land use intensification rather than changes in the area of rice production. Statistics compiled by the International Rice Research Institute²⁶ show that the world's rough rice (that is, unprocessed paddy rice) areas did not change by a substantial amount between 1980 and 2005 (Fig. 3a). Considering rice production areas alone, trends over the past several decades are consistent with a stabilization of fluxes, but not the overall reduction needed to explain the CH_4 mixing ratio and isotope observations presented here.

Accumulating evidence suggests, however, that application of chemical fertilizer⁹ and more efficient water use¹⁰ substantially reduces CH_4 emissions per unit of area of rice production. CH_4 fluxes in rice fields are strongly regulated by levels of organic fertilizer input²⁷. Chemical fertilizer application in China and India increased substantially from 1970 to 2005 (Fig. 3c). The rapid rise in chemical fertilizer

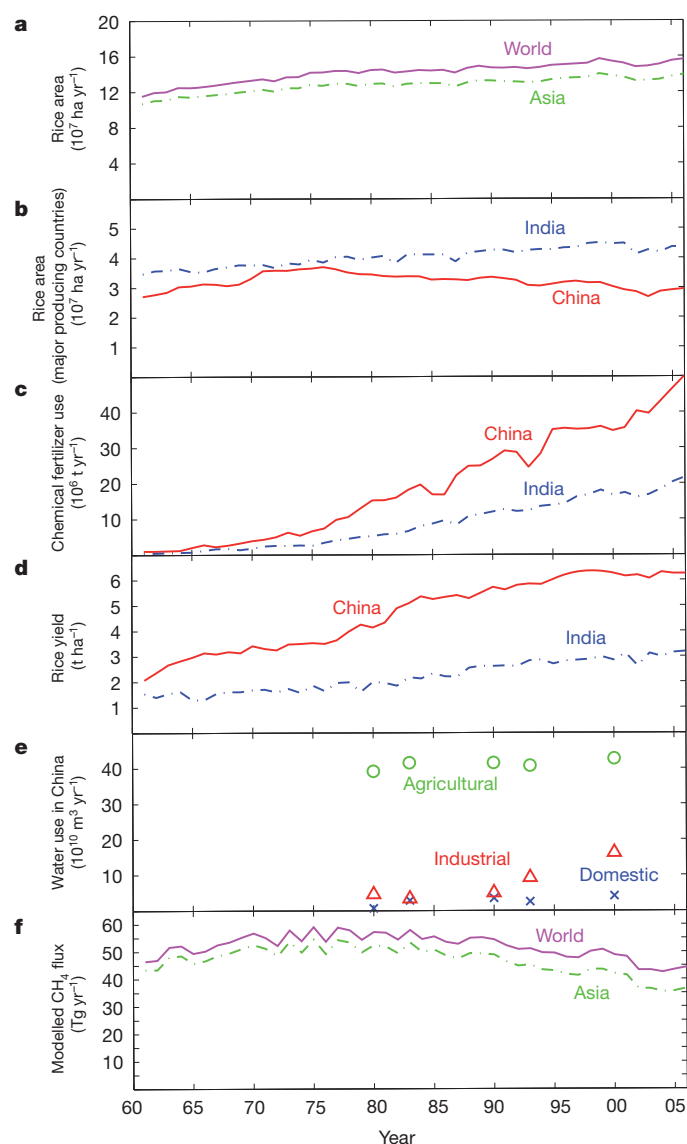


Figure 3 | Evidence for intensification of rice agriculture in Asia. The total world area under rice production has not substantially changed since the late 1970s and is dominated by production in Asian countries (a). The two largest rice-producing countries during 1960–2005 were China and India (b). The use of chemical fertilizer in both countries, especially China, increased substantially during the same time period (c). The increase in rice yield (d) was consistent with rapid increases in chemical fertilizer application and implied a weakening dependence on organic fertilizer, which is both labour-intensive and a primary substrate for CH_4 production. Growing industrial water demand after the 1980s reduced the percentage of water withdrawals available for agriculture in China (e) and may be one factor contributing to new mid-season drainage practices. After the early 1980s, decreases in global CH_4 emissions from rice agriculture were predicted to occur on the basis of an analysis from a biogeochemical process model driven by climate, rice area, fertilizer and water management information (f).

use is consistent with concurrent increases in rice yield (Fig. 3d) and reductions in organic amendments that are more labour-intensive⁹. Other studies have shown that changing the water management of rice paddies can reduce CH_4 emissions by about 10–80% (refs 10, 11 and 27). At the country level, increasing urban and industrial demands for water have increased since 1980 and have limited the water available for agriculture²⁸ (Fig. 3e).

To assess the magnitude of emissions changes associated with rice agriculture during 1960–2005, we modified an empirically based biogeochemical process model^{29,30} to include fertilizer application and

water management impacts on rice yield and methane fluxes (see Supplementary Information). Using this model, which is driven by climate, production area statistics, fertilizer inputs and management information, we estimated that the Northern Hemisphere rice emissions reached a maximum in the early 1980s and decreased by about $15.5 \pm 1.9 \text{ Tgyr}^{-1}$ between 1984 and 2005 (Fig. 3f). The decreasing trend in emissions explained approximately half ($51 \pm 18\%$) of the observed trend inferred from the atmospheric CH_4 measurements (Fig. 2a).

Here we found that the increasing enrichment of $\delta^{13}\text{C}$ in the Northern Hemisphere and a narrowing of the $\delta^{13}\text{C}$ IHD provided new information about the causes of the CH_4 slowdown between 1989 and 2005. Model simulations indicated that decreases in Northern Hemisphere microbial sources had an isotopic signature consistent with the observed atmospheric changes. Although changes in atmospheric OH levels and fossil fuel emissions probably contributed to some of the variability during this period^{5,7,8}, the trends in the $\delta^{13}\text{C}$ IHD presented here excluded these changes as the most important contributors to the long-term CH_4 slowdown. Changes in management of rice agriculture probably contributed to some of the observed decline in Northern Hemisphere emissions, on the basis of estimates from a biogeochemical model. Evaluating the net impact of these management changes on climate requires careful consideration of other greenhouse gas fluxes, including production of N_2O (see, for example, ref. 11). The future trajectory of CH_4 remains difficult to predict, with recent increases after 2006 potentially attributable to emissions from Northern Hemisphere wetlands²⁵ and biomass burning in the tropics³. Our work demonstrates the importance of considering changing agricultural practices in addition to climate effects on natural sources as key drivers of future atmospheric CH_4 levels.

METHODS SUMMARY

To quantify the IHDs of CH_4 , $\delta^{13}\text{C}\text{-CH}_4$ and $\delta\text{D}\text{-CH}_4$, we constructed a long-term time series in each hemisphere (at 40°N and 41°S) by using measurements from the UCI, NIWA and SIL networks^{15–17}. These data were selected because they possessed multi-year $\delta^{13}\text{C}$ and δD measurements of atmospheric CH_4 required in our analysis for constraining CH_4 source and sink processes. We then used a two-box atmospheric model¹⁸ that included information about CH_4 source and sink processes (and their impacts on isotopic fractionation) to examine various hypotheses formulated to explain the overall decline in the CH_4 growth rate. We forced the model using emissions inventories (see, for example, ref. 22) for anthropogenic emissions from 1700 to 1980, and in our control simulation these emissions were held constant thereafter (through 2005). The model included changes in soil and animal CH_4 source isotopic composition over the past two centuries induced by changes in $\delta^{13}\text{C}\text{-CO}_2$. Finally, to evaluate changes in rice agricultural sources over the past few decades, we used an empirical process-based biogeochemical model with parameterizations derived from field measurements^{29,30}. We forced the biogeochemical model with time series of rough rice area and fertilizer application rates using data from the International Rice Research Institute²⁶. Details of the sampling and measuring procedure, comparisons of our measurements with observations from other networks, analysis methods, and modelling approaches are described in the Supplementary Information. The main constraints offered by the observed mixing and isotopic ratios described above are summarized in a schematic figure (Supplementary Fig. 5).

Received 3 March 2010; accepted 18 May 2011.

- Dlugokencky, E. J., Masarie, K. A., Lang, P. M. & Tans, P. P. Continuing decline in the growth rate of the atmospheric methane burden. *Nature* **393**, 447–450 (1998).
- Simpson, I. J., Rowland, F. S., Meinardi, S. & Blake, D. R. Influence of biomass burning during recent fluctuations in the slow growth of global tropospheric methane. *Geophys. Res. Lett.* **33**, L22808 (2006).
- Dlugokencky, E. J. *et al.* Observational constraints on recent increases in the atmospheric CH_4 burden. *Geophys. Res. Lett.* **36**, L18803 (2009).
- Khalil, M. A. K. & Shearer, M. J. in *Atmospheric Methane: its Role in the Global Environment* (ed. Khalil, M. A. K.) 98–111 (Springer, 2000).
- Bousquet, P. *et al.* Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature* **443**, 439–443 (2006).
- Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. D. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
- Fiore, A. M., Horowitz, L. W., Dlugokencky, E. J. & West, J. J. Impact of meteorology and emissions on methane trends, 1990–2004. *Geophys. Res. Lett.* **33**, L12809 (2006).
- Dlugokencky, E. J. *et al.* Atmospheric methane levels off: temporary pause or a new steady-state? *Geophys. Res. Lett.* **30**, 1992, doi:10.1029/2003GL018126 (2003).
- Denier van der Gon, H. Changes in CH_4 emission from rice fields from 1960 to 1990s. 2. The declining use of organic inputs in rice farming. *Glob. Biogeochem. Cycles* **13**, 1053–1062 (1999).
- Li, C. S. *et al.* Reduced methane emissions from large-scale changes in water management of China's rice paddies during 1980–2000. *Geophys. Res. Lett.* **29**, 1972, doi:10.1029/2002GL015370 (2002).
- Frolking, S., Li, C. S., Braswell, R. & Fuglestedt, J. Short- and long-term greenhouse gas and radiative forcing impacts of changing water management in Asian rice paddies. *Glob. Change Biol.* **10**, 1180–1196 (2004).
- Worthy, D. E. J. *et al.* Decreasing anthropogenic methane emissions in Europe and Siberia inferred from continuous carbon dioxide and methane observations at Alert, Canada. *J. Geophys. Res.* **114**, D10301, doi:10.1029/2008JD011239 (2009).
- Cicerone, R. J. & Oremland, R. S. Biogeochemical aspects of atmospheric methane. *Glob. Biogeochem. Cycles* **2**, 299–327 (1988).
- Lassey, K. R., Etheridge, D. M., Lowe, D. C., Smith, A. M. & Ferretti, D. F. Centennial evolution of the atmospheric methane budget: what do the carbon isotopes tell us? *Atmos. Chem. Phys.* **7**, 2119–2139 (2007).
- Tyler, S. C., Rice, A. L. & Ajie, H. O. Stable isotope ratios in atmospheric CH_4 : implications for seasonal sources and sinks. *J. Geophys. Res.* **112**, D03303 (2007).
- Lowe, D. C., Manning, M. R., Brailsford, G. W. & Bromley, A. M. The 1991–1992 atmospheric methane anomaly: Southern Hemisphere ^{13}C decrease and growth rate fluctuations. *Geophys. Res. Lett.* **24**, 857–860 (1997).
- Quay, P. *et al.* The isotopic composition of atmospheric methane. *Glob. Biogeochem. Cycles* **13**, 445–461 (1999).
- Tans, P. P. A note on isotopic ratios and the global atmospheric methane budget. *Glob. Biogeochem. Cycles* **11**, 77–81 (1997).
- Etheridge, D. M., Steele, L. P., Francey, R. J. & Langenfelds, R. L. Atmospheric methane between 1000 AD and present: evidence of anthropogenic emissions and climatic variability. *J. Geophys. Res.* **103**, 15979–15993 (1998).
- Ferretti, D. F. *et al.* Unexpected changes to the global methane budget over the past 2000 years. *Science* **309**, 1714–1717 (2005).
- Lassey, K. R. Livestock methane emission: from the individual grazing animal through national inventories to the global methane cycle. *Agric. For. Meteorol.* **142**, 120–132 (2007).
- van Aardenne, J. A., Dentener, F. J., Olivier, J. G. J., Goldewijk, C. G. M. K. & Lelieveld, J. A $1^\circ \times 1^\circ$ resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990. *Glob. Biogeochem. Cycles* **15**, 909–928 (2001).
- Global Anthropogenic Non- CO_2 Greenhouse Gas Emissions: 1990–2020 (<http://www.epa.gov/climatechange/economics/downloads/GlobalAnthroEmissionsReport.pdf>) (United States Environmental Protection Agency, 2006).
- Zhuang, Q. *et al.* Methane fluxes between terrestrial ecosystems and the atmosphere at northern high latitudes during the past century: a retrospective analysis with a process-based biogeochemistry model. *Glob. Biogeochem. Cycles* **18**, GB3010 (2004).
- Bloom, A. A., Palmer, P. I., Fraser, A., Reay, D. & Frankenberg, C. Large-scale controls of methanogenesis inferred from methane and gravity spaceborne data. *Science* **327**, 322–325 (2010).
- World Rice Statistics (<http://irri.org/world-rice-statistics>) (International Rice Research Institute, 2008; accessed 22 September 2008).
- Wassmann, R. *et al.* Characterization of methane emissions from rice fields in Asia. III. Mitigation options and future research needs. *Nutr. Cycl. Agroecosyst.* **58**, 23–36 (2000).
- Water Use Statistics (<http://www.fao.org/nr/water/aquastat/main/index.stm>) (Food and Agriculture Organization of the United Nations, 2008; accessed 22 September 2008).
- Huang, Y., Sass, R. L. & Fisher, F. M. A semi-empirical model of methane emission from flooded rice paddy soils. *Glob. Change Biol.* **4**, 247–268 (1998).
- Kai, F. M., Tyler, S. C. & Randerson, J. T. Modeling methane emissions from rice agriculture in China during 1961–2007. *J. Integr. Env. Sci.* **7**, 49–60 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the many researchers associated with the laboratories referred to as NOAA, NIWA, UW and UCI who took part in collecting and measuring many thousands of air samples over the years. We also thank P. Bousquet for detailed comments on earlier drafts that improved the manuscript. This work has been funded by NASA grants to S.C.T. (NGT5-30409) and J.T.R. (NNX08AF64G), NSF grants to S.C.T. (ATM 9871077) and J.T.R. (ATM-0628637 and AGS-1021776), and additional support from the W. M. Keck Foundation.

Author Contributions S.C.T., F.M.K. and D.R.B. carried out CH_4 , $\delta^{13}\text{C}\text{-CH}_4$ and $\delta\text{D}\text{-CH}_4$ measurements. F.M.K., S.C.T. and J.T.R. designed the study. F.M.K. conducted data analysis and performed the model simulations. F.M.K., S.C.T. and J.T.R. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.M.K. (fmkai@smart.mit.edu).

Recent decreases in fossil-fuel emissions of ethane and methane derived from firn air

Murat Aydin¹, Kristal R. Verhulst¹, Eric S. Saltzman¹, Mark O. Battle², Stephen A. Montzka³, Donald R. Blake¹, Qi Tang¹ & Michael J. Prather¹

Methane and ethane are the most abundant hydrocarbons in the atmosphere and they affect both atmospheric chemistry and climate. Both gases are emitted from fossil fuels and biomass burning, whereas methane (CH_4) alone has large sources from wetlands, agriculture, landfills and waste water. Here we use measurements in firn (perennial snowpack) air from Greenland and Antarctica to reconstruct the atmospheric variability of ethane (C_2H_6) during the twentieth century. Ethane levels rose from early in the century until the 1980s, when the trend reversed, with a period of decline over the next 20 years. We find that this variability was primarily driven by changes in ethane emissions from fossil fuels; these emissions peaked in the 1960s and 1970s at 14–16 teragrams per year ($1 \text{ Tg} = 10^{12} \text{ g}$) and dropped to 8–10 Tg yr^{-1} by the turn of the century. The reduction in fossil-fuel sources is probably related to changes in light hydrocarbon emissions associated with petroleum production and use. The ethane-based fossil-fuel emission history is strikingly different from bottom-up estimates of methane emissions from fossil-fuel use^{1,2}, and implies that the fossil-fuel source of methane started to decline in the 1980s and probably caused the late twentieth century slow-down in the growth rate of atmospheric methane^{3,4}.

Ethane is an organic trace gas that is primarily emitted to the atmosphere during mining, processing, transport and consumption of fossil fuels, during use of biofuels, and during biomass burning^{4–6}. It acts as a precursor of ozone and carbon monoxide in the troposphere. Ethane is oxidized rapidly by the hydroxyl radical (OH^\bullet) and has a seasonally varying lifetime (annual mean ~ 2 months). The short lifetime, coupled with a north–south asymmetry in its sources, leads to geographic and temporal variability in its atmospheric abundance^{5,7}. Ethane abundance displays strong seasonal variability, with peak-to-peak amplitudes comparable to annual-mean levels. Annual-mean ethane levels are highest at high northern latitudes (HNL, that is, $30^\circ\text{--}90^\circ \text{ N}$; 1.5 parts per billion) and lowest at high southern latitudes (HSL, that is, $30^\circ\text{--}90^\circ \text{ S}$; 250 parts per trillion)⁴.

We measured ethane in firn air collected at Summit (in Greenland), at the West Antarctic Ice Sheet – Divide, WAIS-D (in Antarctica), and at the South Pole. A synthesis inversion method was used to develop atmospheric histories of ethane for each site, using a one-dimensional firn air diffusion model (Fig. 1) (see Methods Summary). The mean temperatures and the ice accumulation rates are very similar at Summit (-31° C , 20 cm yr^{-1}) and at WAIS-D (-31° C , 22 cm yr^{-1}), resulting in similar gas age distributions in the firn and atmospheric histories constrained for the past 50–60 years (refs 8, 9). South Pole is considerably colder (-51° C), with a lower ice accumulation rate ($\sim 8 \text{ cm yr}^{-1}$) and a deeper firn column. As a result, atmospheric histories based on South Pole firn air measurements are constrained for the past 80–90 years (ref. 10). The ethane atmospheric histories based on the WAIS-D and South Pole firn air measurements display the same trends for the period since 1950 (Fig. 1).

Mean annual ethane levels measured over Summit and South Pole are consistent with those at other high-latitude sites, indicating that the

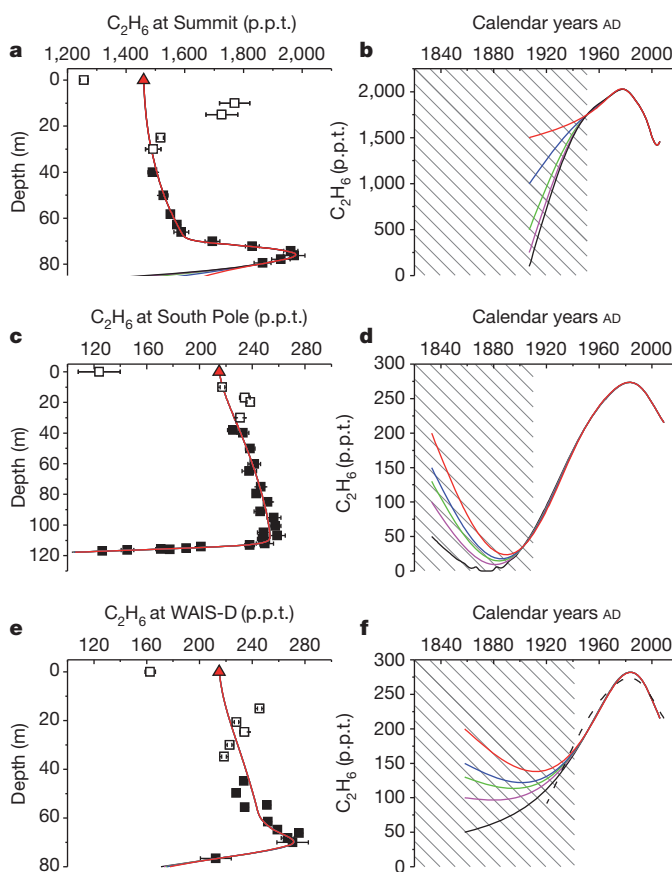


Figure 1 | Ethane mixing ratios in firn air at three sites, and the atmospheric histories derived from these measurements. **a–f**, Results for Summit (**a**, **b**), South Pole (**c**, **d**) and WAIS-D (**e**, **f**). Filled and open squares in **a**, **c** and **e** show measurements with estimated uncertainties (error bars, $\pm 2 \text{ s.e.}$) (Supplementary Data). Solid lines in **a**, **c** and **e** show modelled firn profiles for the three respective sites, with the five different atmospheric histories (solid lines) shown in **b**, **d** and **f**. These five atmospheric histories were obtained by inverse modelling of the measurements with five different boundary conditions at t_0 that are identified by different colours, each representing a different pre-industrial mixing ratio for ethane. The data points in the top 35 m (Summit and South Pole) and 40 m (WAIS-D) of the firn are subject to the effects of seasonal variations in surface ethane levels (Supplementary Information) and are ignored during inverse modelling (open squares in **a**, **c** and **e**). The inversions were forced with contemporary annual-mean surface mixing ratios of 1,460 p.p.t. at Summit and 215 p.p.t. at South Pole and WAIS-D (red triangles in **a**, **c** and **e**) (Supplementary Information). The South Pole atmospheric history is overlaid on the independently derived WAIS-D atmospheric histories for comparison (dashed line in **f**). The inversions are sensitive to assumptions about the pre-industrial ethane levels before 1950, 1940 and 1910 at Summit, WAIS-D and South Pole, respectively (shaded areas), implying that only the inversion results for the later years are valid atmospheric histories.

¹Department of Earth System Science, University of California, Irvine, California 92697, USA. ²Department of Physics and Astronomy, Bowdoin College, Brunswick, Massachusetts 04011, USA. ³Earth System Research Laboratories – Global Monitoring Division, National Oceanic and Atmospheric Administration, Boulder, Colorado 80305, USA.

measurements in polar firn air provide a reasonable sampling of the ethane mixing ratios in high latitudes (see Supplementary Information). Together, the three atmospheric histories show ethane levels peaking in the HNL and HSL atmospheres in the early 1980s, followed by a decline over the 20 years that followed. The South Pole site contains sufficiently old air to constrain the ramp-up period during the early to mid twentieth century (Fig. 1). On the basis of these atmospheric histories, HSL annual-mean ethane levels increased fivefold between 1910 and 1980 (from about 60 p.p.t. to 280 p.p.t.), and declined more than 10% (to less than 250 p.p.t.) by 2000. In the HNL, ethane increased from ~ 1.7 p.p.b. in 1950 to ~ 2.0 p.p.b. in 1980, and declined 25% to ~ 1.5 p.p.b. in 2000.

To infer the causes of the atmospheric ethane variations in terms of changes in large-scale sources, we must relate high-latitude ethane levels to hemispheric averages. We used the modern atmospheric ethane distribution^{4,6} and sensitivity tests with the UCI Chemical Transport Model (UCI-CTM)^{11,12} to derive ratios relating the response of the HNL and HSL to changes in hemispheric mean ethane levels (see Supplementary Information). Next, we used a simple two-box model (see Methods Summary) representing the troposphere in the Northern and Southern hemispheres to simulate variations in mean atmospheric ethane levels over the past century resulting from various scenarios for fossil-fuel, biomass-burning, and biofuel emissions. Fossil-fuel and biofuel emissions are concentrated in the northern mid-latitudes, and biomass burning in the tropics. As a result, these sources contribute to hemispheric ethane levels with different efficiencies (see Methods Summary and Supplementary Information).

The two-box model was used to infer the fossil-fuel ethane emission histories needed to achieve agreement with the firn-air-derived atmospheric ethane histories from Summit and South Pole, under various assumptions about biomass-burning emissions (Fig. 2, see Supplementary Information). The results show clearly that ethane variability over decadal timescales during the twentieth century was dominated by changes in the fossil-fuel ethane source. When biomass-burning and fossil-fuel ethane emissions are both allowed to vary after 1950, the model yields a biomass-burning source of less than 1 Tg yr^{-1} in 1950, rising to $\sim 3 \text{ Tg yr}^{-1}$ by 2000, followed by a drop to 2 Tg yr^{-1} . The model results show a large change in fossil-fuel ethane emissions in the first half of the twentieth century, peaking at $14\text{--}16 \text{ Tg yr}^{-1}$ in the late 1950s. Fossil-fuel emissions are then constant for about 20 years, followed by a 45% drop over the next 30 years. The model yields $8\text{--}10 \text{ Tg yr}^{-1}$ for fossil-fuel ethane emissions at the end of the twentieth century, consistent with the modern ethane budget inferred from atmospheric ethane levels^{6,13}.

Fossil-fuel emissions are also a significant source of methane into the atmosphere, and the observed decline in the growth rate of atmospheric methane^{3,4} parallels the decline in atmospheric ethane levels during 1980–2000. This evidence suggests that the decline in methane growth rates was caused by a gradual reduction in fossil-fuel emissions, which was already underway when continuous direct atmospheric measurements of methane started in 1984 and 1985^{3,4}. The stabilization of atmospheric ethane in recent years⁴ (also see Supplementary Information) suggests that fossil-fuel emissions are now steady. Thus, the recent increases in atmospheric methane levels are probably not derived from increased fossil-fuel emissions and must be due to other sources, as previously suggested^{14,15}.

The ethane-based biomass-burning emission history agrees well with some independent estimates^{6,16}. However, our fossil-fuel emissions history is quite different from bottom-up methane emission inventories^{1,2} (Fig. 3). Most notably, the ethane-based fossil-fuel emissions display a steep ramp-up after 1920 and a sharp decline after 1980, whereas the bottom-up methane fossil-fuel emission inventories display a generally increasing trend through the entire twentieth century, with the sharpest increases occurring between 1950 and 1980. Hydrocarbons are emitted from a variety of fossil-fuel sources, each exhibiting a different range of methane/ethane emission ratio (MER). For example,

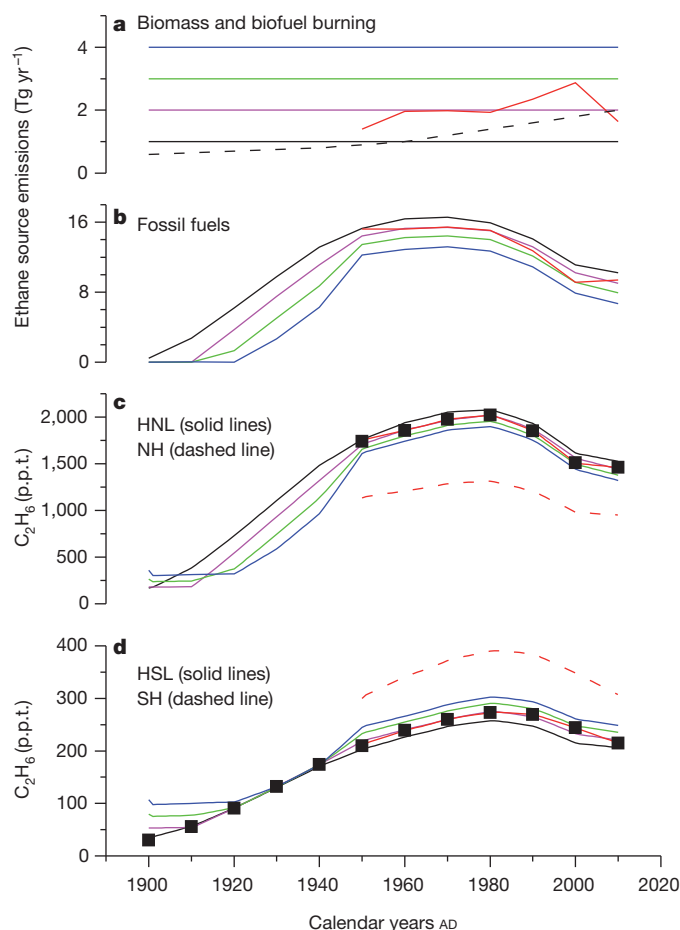


Figure 2 | Ethane source emissions and the resulting atmospheric histories.

a–d, Historical global ethane emissions (**a**, **b**) and the resulting atmospheric histories (**c**, **d**) are derived with the two-box model to be consistent with the site-specific atmospheric histories based on firn-air results (Fig. 1). Fossil-fuel emission histories were developed by considering five different biomass-burning cases. In the first four cases, biomass-burning emissions are fixed at 1, 2, 3 and 4 Tg yr^{-1} (black, purple, green and blue lines, respectively; **a–d**), and the fossil-fuel emissions are varied to minimize the χ^2 value of the fit to the firn-air-based ethane atmospheric histories (black squares in **c** and **d**) from both the high northern latitudes and high southern latitudes (HNL and HSL) during 1950–2000, and from HSL only during 1900–1940. We included 2010 in the box model optimizations by assuming the atmospheric ethane levels remained constant at the mixing ratios used as surface tie-points used in firn inversions (Fig. 1). As a fifth case, the fit to the atmospheric histories is optimized by allowing both the fossil-fuel and the biomass-burning emissions to vary (thick red lines). Variable biomass burning is considered only for the period since 1950, because atmospheric histories from both hemispheres are needed to constrain the partitioning between fossil-fuel and biomass-burning emissions. Emissions from biofuel burning are fixed at the historical estimates (dashed black line; **a**) in all five cases². The atmospheric lifetime of ethane (Methods Summary) is also the same in all five cases. The hemispheric-average atmospheric histories are also shown for the fifth case (dashed red lines; **c** and **d**), in which fossil-fuel and biomass-burning emissions are both varied. Ethane levels in HNL and HSL are obtained from Northern and Southern hemispheric (NH and SH) levels using the following ratios for fossil-fuel, biofuels and biomass-burning emissions: HNL/NH: 1.6, 1.4, 1.0; HSL/SH: 0.7, 0.7, 0.7. Values differ from 1.0 owing to the latitudinal distribution of each type of emission (Supplementary Information).

coal and natural gas emissions are methane-rich ($\text{MER} = 5\text{--}100$) with respect to emission ratios measured near oil-storage and oil-processing facilities ($\text{MER} = 3\text{--}5$)⁶. It is possible that the global-average MER (the ratio of total methane to total ethane emissions from all fossil-fuel sources combined) changed over time as the relative emission strengths of the different fossil-fuel sources varied during the twentieth century,

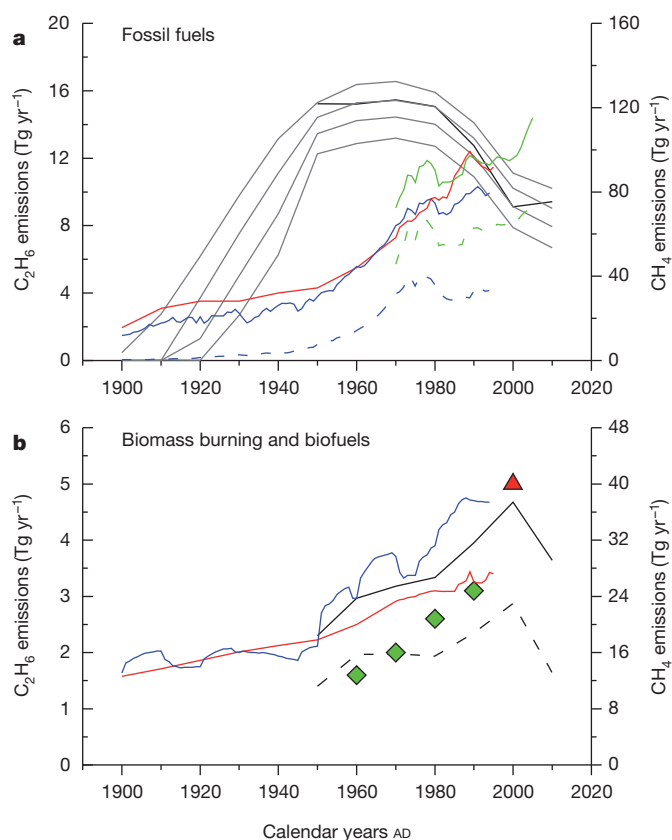


Figure 3 | Ethane and methane emissions from fossil fuels, biofuels and biomass burning. **a**, Ethane-based fossil-fuel emission histories (left y-axis) for fixed biomass burning (solid grey lines) and variable biomass burning (solid black line) cases, compared with bottom-up methane emission estimates from fossil fuels (coloured lines, right y-axis) based on the following sums of emissions: blue, gas flaring, gas supply and coal mining¹; red, fossil-fuel consumption, fossil-fuel production and industrial²; green, oil, gas and solid fuels (data from <http://edgar.jrc.ec.europa.eu/>). The oil and gas only methane emissions from EDGAR4.1 (see URL in previous sentence; green dashed line) and gas flaring and gas supply methane emissions from ref. 1 (blue dashed line) are also shown. The trends in the ethane-based histories are relatively insensitive to assumptions about biomass burning. None of the bottom-up methane fossil-fuel emission inventories display a persistent decline during 1980–2000 that would be consistent with the ethane-based emission histories. The scales of the left and right y-axes are set at a ratio of 8 because it suits the presentation well and is consistent with **b**. **b**, Ethane emissions (left y-axis) from biomass burning (dashed black line) and biomass-burning+biofuels (solid black line) for the variable biomass-burning case in the two-box model compared with methane emissions (right y-axis) attributed to the total burning product from ref. 1 (blue line)¹ and with EDGAR-HYDE 1.4 (red line)². The scales of the left and right y-axes are set at a ratio of 8, which is a measurement-based estimate of methane/ethane emission ratio (MER) from biomass burning³⁰. Our results and the bottom-up methane inventories are consistent in how hydrocarbon emissions from biomass burning changed over time and suggest an MER of 8 on a Tg per Tg basis, in agreement with the earlier estimates³⁰. Our estimate of ethane biomass-burning emissions from the variable biomass-burning case is also in good agreement with some independent estimates of historical ethane emissions from biomass burning¹⁶ (green diamonds, left y-axis) and a total burning emissions estimate (biomass burning+biofuels) for present day⁶ (red triangle, left y-axis).

but the apparent differences between the ethane-based fossil-fuel emission histories and inventory-based methane emissions (Fig. 3) cannot be reconciled with such considerations. Inconsistencies are already apparent in mid-century, when the ratio of bottom-up methane emissions to ethane-based fossil-fuel emissions imply an MER of 2–3 (Fig. 3). It is highly unlikely that the global-average MER during the 1950s could be lower than point sources with the lowest MERs, especially considering that both coal and natural gas were sources of significant hydrocarbon

emissions during that period. This suggests the bottom-up methane inventories underestimate fossil-fuel emissions at mid-century.

The ethane-based estimates of fossil-fuel emissions show a decline during the 1980s, while the methane-inventory-based estimates of fossil-fuel emissions show an increase (Fig. 3). Opposing trends in the fossil-fuel emissions of ethane and methane are highly unlikely, because this requires ethane to be selectively removed from both newly introduced and existing hydrocarbon sources while the residual methane is released to the atmosphere. Natural gas is used in the production of feedstock ethane in plastics manufacturing. However, it seems implausible that methane, which is economically the most valuable component of natural gas, would simply be vented back into the atmosphere after ethane had been removed from natural gas. In fact, the amount of feedstock ethane produced from natural gas is inversely related to the price of natural gas¹⁷, and probably declined as natural gas became progressively more valuable during the second half of the twentieth century. We conclude that the discrepancy between the ethane-based fossil-fuel emission histories and the fossil-fuel emission histories from the bottom-up methane inventories cannot be explained by realistic changes in global-average MER during the twentieth century.

The decline in fossil-fuel emissions, as calculated from our firm-air ethane records, coincides with a period of rapid expansion in natural gas production during the second half of the twentieth century^{18,19}. We speculate that the rising economic value of natural gas during the late twentieth century¹⁹ and the development of cleaner technologies led to sharp reductions in the release of light hydrocarbons into the atmosphere. Emissions linked specifically to the growing natural gas industry must have been more than offset by large reductions in the venting of light hydrocarbons, including methane and ethane, associated with production and processing of petroleum. These changes appear to be underestimated in the bottom-up methane inventories. We estimate that the total decline in fossil-fuel emissions of ethane was 5–6 Tg yr^{−1} during 1980–2000. Attributing this decline entirely to decreases in fossil-fuel emission sources with the lowest MER (3–5) implies a 15–30 Tg yr^{−1} drop in fossil-fuel emissions of methane. Distributing the decline among different fossil-fuel sources would yield a larger change in methane emissions from fossil fuels. An independent inversion analysis of methane observations suggests that a 20 Tg yr^{−1} drop in fossil-fuel methane emissions during the 1990s contributed to the decline in methane growth rates²⁰. Our results are consistent with such a decline, but provide additional evidence that the drop in fossil-fuel emissions started sooner, by about a decade. The decline in fossil-fuel emissions during the 1990s accounts for less than 60% of the total reduction during 1980–2000 in the two-box model results for ethane, suggesting that the reduction in fossil-fuel methane emissions could be ~40% larger than the previous estimate²⁰ of 20 Tg yr^{−1}.

Given that we do not have comparable observations from lower latitudes, it is possible that a shift in the location of fossil-fuel emissions towards lower latitudes within the Northern Hemisphere contributed to the observed ethane decline during 1980–2000. However, we estimate this effect to be relatively small on the basis of sensitivity tests with the UCI-CTM, which show that possible changes in the location of the emission reductions for ethane could add about 20% to the magnitude of emission reductions calculated above (see Supplementary Information).

It is also possible that the late twentieth century decline in ethane could have been caused by a decrease in the atmospheric lifetime of ethane, but we estimate the likelihood for such changes to be low. Assuming fixed emissions, a 5–6 Tg yr^{−1} increase in ethane loss (roughly one-third of the peak budget) would be required to account for the observed ethane decline. A change of this magnitude due to increases in OH[•] concentrations during the late twentieth century is unlikely^{21,22}; however, atmospheric levels of chlorine atoms (Cl[•]) might have increased, because of increasing tropospheric NO_x, increasing tropospheric ozone and increasing acidity of aerosols^{23,24}. A Cl[•] sink

of the magnitude required to explain the ethane decline would have a minor effect on the methane budget ($\sim 3\%$), because the relative reactivity of Cl^\bullet versus OH^\bullet ($k_{\text{Cl}}/k_{\text{OH}}$) is considerably smaller for methane than it is for ethane. Accounting for the ethane decline would require an increase of roughly $1.1 \times 10^4 \text{ Cl}^\bullet \text{ cm}^{-3}$ over the entire marine boundary layer, or a much larger increase if just the polluted boundary layer is considered. Estimates for the clean marine boundary layer are in the range $10^3\text{--}10^4 \text{ Cl}^\bullet \text{ cm}^{-3}$. It is not possible to assess whether a change of the required magnitude is viable with the limited observational data available at present^{23,25}.

METHODS SUMMARY

Firn-air measurements, modelling and inversion. At all three sites, multiple flasks (see Supplementary Information) from the same depth were filled sequentially, using established methods¹⁰. All flasks were analysed at the University of California, Irvine, using $\sim 100 \text{ cm}^3$ STP samples on a gas chromatography-mass spectrometry (GC-MS) system designed for trace gas analysis on small air samples²⁶. Summit and South Pole flasks were analysed at least twice. We used a one-dimensional firn-air model and a synthesis inversion method to derive atmospheric histories of the annual-mean, high-latitude, tropospheric abundances of ethane from the firn-air measurements (see Supplementary Information).

Two-box model. The Northern and Southern hemispheres are represented as equal mass boxes of $2.2 \times 10^9 \text{ Tg}$ (exchange time 1 year). Ethane is lost through OH^\bullet oxidation ($1/2.6 \text{ months}^{-1}$) and transport into the stratosphere ($1/35 \text{ yr}^{-1}$), yielding a present day lifetime of 2.3 months. The annual-mean OH^\bullet loss frequency varies over the past 100 years based on the methane feedback on tropospheric OH^\bullet : -0.32% in OH^\bullet for every $+1\%$ in methane²⁷. The OH^\bullet changes differently in Northern and Southern hemispheres: -0.305% and -0.335% , respectively, based on simulations with the UCI-CTM^{11,12}. We adopt a methane increase of 900 p.p.b. to 1,790 p.p.b. since 1900²⁸. Fossil-fuel, biofuel and biomass-burning sources in the model Northern Hemisphere are adopted from previous work⁶ and represent 93%, 81% and 58%, respectively, of the total emissions. The fossil-fuel emission histories for various biomass-burning scenarios are developed by an inverse optimization algorithm (see Supplementary Information).

Three-dimensional model. The UCI-CTM's simulation of ethane with realistic distribution of sources²⁹ is used to relate the northern and southern tropospheric mean abundances to the annual-mean abundance over the ice sheets, and thus provide a correction to the two-box model histories for the firn-air modelling.

Received 15 September 2010; accepted 6 July 2011.

- Stern, D. I. & Kaufmann, R. K. Estimates of global anthropogenic methane emissions 1860–1993. *Chemosphere* **33**, 159–176 (1996).
- van Aardenne, J. A., Dentener, F. J., Olivier, J. G. J., Klein Goldewijk, C. G. M. & Lelieveld, J. A. $1^\circ \times 1^\circ$ resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990. *Glob. Biogeochem. Cycles* **15**, 909–928 (2001).
- Dlugokencky, E. J. *et al.* Atmospheric methane levels off: temporary pause or a new steady-state? *Geophys. Res. Lett.* **30**, 1992, doi:10.1029/2003GL018126 (2003).
- Simpson, J. J., Rowland, F. S., Meinardi, S. & Blake, D. R. Influence of biomass burning during recent fluctuations in the slow growth of global tropospheric methane. *Geophys. Res. Lett.* **33**, L22808, doi:10.1029/2006GL027330 (2006).
- Rudolph, J. The tropospheric distribution and budget of ethane. *J. Geophys. Res.* **100**, 11369–11381 (1995).
- Xiao, Y. *et al.* Global budget of ethane and regional constraints on U.S. sources. *J. Geophys. Res.* **113**, D21306, doi:10.1029/2007JD009415 (2008).
- Boissard, C., Bonsang, B., Kanakidou, M. & Lambert, G. TROPZ II: global distributions and budgets of methane and light hydrocarbons. *J. Atmos. Chem.* **25**, 115–148 (1996).
- Aydin, M. *et al.* Post-coring entrapment of modern air in some shallow ice cores collected near the firn-ice transition: evidence from CFC-12 measurements in Antarctic firn air and ice cores. *Atmos. Chem. Phys.* **10**, 5135–5144 (2010).
- Fain, X. *et al.* Mercury in the snow and firn at Summit Station, Central Greenland, and implications for the study of past atmospheric mercury levels. *Atmos. Chem. Phys.* **8**, 3441–3457 (2008).
- Battle, M. *et al.* Atmospheric gas concentrations over the past century measured in air from firn at the South Pole. *Nature* **383**, 231–235 (1996).
- Hsu, J., Prather, M. J. & Wild, O. Diagnosing the stratosphere-to-troposphere flux of ozone in a chemistry transport model. *J. Geophys. Res.* **110**, D19305, doi:10.1029/2005JD006045 (2005).
- Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D. & Rodriguez, J. M. Quantifying errors in trace gas species transport modeling. *Proc. Natl Acad. Sci. USA* **105**, 19617–19621 (2008).
- Pozzer, A. *et al.* Observed and simulated global distribution and budget of atmospheric $\text{C}_2\text{--C}_6$ alkanes. *Atmos. Chem. Phys.* **10**, 4403–4422 (2010).
- Dlugokencky, E. J. *et al.* Observational constraints on recent increases in the atmospheric CH_4 burden. *Geophys. Res. Lett.* **36**, L18803, doi:10.1029/2009GL039780 (2009).
- Rigby, M. *et al.* Renewed growth of atmospheric methane. *Geophys. Res. Lett.* **35**, L22805, doi:10.1029/2008GL036037 (2008).
- Schultz, M. G. *et al.* Global wildland fire emissions from 1960 to 2000. *Glob. Biogeochem. Cycles* **22**, GB2002, doi:10.1029/2007GB003031 (2008).
- Remer, D. S. & Jorgens, C. Ethylene economics and production forecasting in a changing environment. *Eng. Process Econ.* **3**, 267–278 (1978).
- Barns, D. W. & Edmonds, J. A. *An Evaluation of the Relationship Between the Production and Use of Energy and Atmospheric Methane Emissions* (TR047, DOE/NBB-0088P, National Technical Information Service, US Dept of Commerce, Springfield, 1990).
- Gillardoni, A. *The World Market for Natural Gas; Implications for Europe* (Springer, 2008).
- Bousquet, P. *et al.* Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature* **443**, 439–443 (2006).
- Prinn, R. G. *et al.* Evidence for variability of atmospheric hydroxyl radicals over the past quarter century. *Geophys. Res. Lett.* **32**, L07809, doi:10.1029/2004GL022228 (2005).
- Montzka, S. A. *et al.* Small interannual variability of global atmospheric hydroxyl. *Science* **331**, 67–69 (2011).
- Lawler, M. J. *et al.* Pollution-enhanced reactive chlorine chemistry in the eastern tropical Atlantic boundary layer. *Geophys. Res. Lett.* **36**, L08810, doi:10.1029/2008GL036666 (2009).
- Thornton, J. A. *et al.* A large atomic chlorine source inferred from mid-continental reactive nitrogen chemistry. *Nature* **464**, 271–274 (2010).
- Allan, W., Struthers, H. & Lowe, D. C. Methane carbon isotope effects caused by atomic chlorine in the marine boundary layer: global model results compared with southern hemisphere measurements. *J. Geophys. Res.* **112**, D04306, doi:10.1029/2006JD007369 (2007).
- Aydin, M., Williams, M. B. & Saltzman, E. S. Feasibility of reconstructing paleoatmospheric records of selected alkanes, methyl halides, and sulfur gases from Greenland ice cores. *J. Geophys. Res.* **112**, D07312, doi:10.1029/2006JD008027 (2007).
- Prather, M. *et al.* in *Climate Change 2001: The Scientific Basis* (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, 2001).
- MacFarling Meure, C. *et al.* Law Dome CO_2 , CH_4 , and N_2O ice core records extended to 2000 years BP. *Geophys. Res. Lett.* **33**, L14810, doi:10.1029/2006GL026152 (2006).
- Tang, Q. & Prather, M. J. Correlating tropospheric column ozone with tropopause folds: the Aura-OMI satellite data. *Atmos. Chem. Phys.* **10**, 9681–9688 (2010).
- Andreae, M. O. & Merlet, P. Emission of trace gases and aerosols from biomass burning. *Glob. Biogeochem. Cycles* **15**, 955–966 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Sowers, M. Drier and ICDS drillers for support during firn-air sampling and drilling in the field, and T. Sowers for discussions during preparation of the manuscript. We thank M. Bender and J. Severinghaus for ^{15}N measurements in firn air, S. Meinardi for ethane measurements in surface flasks, and P. Tans and P. Lang for firn-air CO_2 data. This work was supported by the National Science Foundation (grants ANT-0739598, ANT-0440602, ANT-0440509, ARC-0520460) and NASA (grant NAG58935).

Author Contributions M.A.: firn-air sampling, ethane analysis in firn air and surface-air flasks, firn-air modelling, two-box modelling, box-model inversions, manuscript preparation. K.R.V.: ethane analysis in firn air and surface-air flasks, firn-air modelling, two-box modelling, firn-air and two-box model inversions, manuscript improvements. E.S.S.: firn-air modelling, two-box modelling, firn-air and two-box model inversions, manuscript improvements. M.O.B.: firn-air sampling, firn-air modelling, manuscript improvements. S.A.M.: halocarbon measurements in firn air to constrain firn processes, NOAA surface air samples, manuscript improvements. D.R.B.: ethane measurements in surface air, manuscript improvements. Q.T.: CTM modelling, manuscript improvements. M.J.P.: CTM modelling, manuscript improvements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A. (maydin@uci.edu).

C₄ grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland

Jack A. Morgan¹, Daniel R. LeCain¹, Elise Pendall², Dana M. Blumenthal¹, Bruce A. Kimball³, Yolima Carrillo², David G. Williams⁴, Jana Heisler-White⁴, Feike A. Dijkstra^{1,5} & Mark West¹

Global warming is predicted to induce desiccation in many world regions through increases in evaporative demand^{1–3}. Rising CO₂ may counter that trend by improving plant water-use efficiency^{4,5}. However, it is not clear how important this CO₂-enhanced water use efficiency might be in offsetting warming-induced desiccation because higher CO₂ also leads to higher plant biomass, and therefore greater transpirational surface^{2,6,7}. Furthermore, although warming is predicted to favour warm-season, C₄ grasses, rising CO₂ should favour C₃, or cool-season plants⁸. Here we show in a semi-arid grassland that elevated CO₂ can completely reverse the desiccating effects of moderate warming. Although enrichment of air to 600 p.p.m.v. CO₂ increased soil water content (SWC), 1.5/3.0 °C day/night warming resulted in desiccation, such that combined CO₂ enrichment and warming had no effect on SWC relative to control plots. As predicted, elevated CO₂ favoured C₃ grasses and enhanced stand productivity, whereas warming favoured C₄ grasses. Combined warming and CO₂ enrichment stimulated above-ground growth of C₄ grasses in 2 of 3 years when soil moisture most limited plant productivity. The results indicate that in a warmer, CO₂-enriched world, both SWC and productivity in semi-arid grasslands may be higher than previously expected.

Grass-dominated, dry rangelands account for over 30% of Earth's terrestrial surface^{9,10} and provide most of the forage for the world's domestic livestock. Among the most important of these include the vast Central Asian steppes, North American prairies, Australian rangelands, plus extensive grazing lands of Africa, South America and the Mediterranean. Productivity of these lands, which is under increasing pressure with human population growth, is primarily limited by water^{9,11}. Soil water availability is driven by complex interactions between precipitation, temperature^{1,11}, ambient CO₂ (ref. 4) and soil properties, rendering rangelands and the livestock and cultures they support particularly susceptible to climate change and rising CO₂. Higher CO₂ concentrations induce stomatal closure, which reduces leaf transpiration and increases plant water-use efficiency and SWC⁴. Despite agreement among global climate models that climate change will induce drought-like conditions in several rangeland-dominated world regions^{1,2}, considerable uncertainty remains as to how stomatal closure from CO₂ might interact with increased temperature to influence stand-level evapotranspiration, soil water dynamics and plant productivity. Many believe that CO₂-induced reductions in transpiration at the leaf level will be largely offset at the canopy level by increases in leaf area^{2,6,7,12}.

Most rangelands are dominated by graminoids, and contain a mixture of the two major photosynthetic pathway classes of plants, C₃ and C₄. Today, C₄ grasses are most common in grasslands and savannas within 40° of the Equator⁸, and are predicted to become more abundant with increasing temperature¹³. However, C₃ plants, owing to their CO₂-limited photosynthetic metabolism, may become more competitive in future CO₂-enriched rangelands^{8,14}. The relative abundances

of these two photosynthetic types govern spatial and temporal aspects of water and nutrient cycling, net primary production and plant-animal interactions^{13,15,16}, yet there is considerable uncertainty about which group will be favoured under future, warmer, CO₂-enriched conditions.

In the Prairie Heating and CO₂ Enrichment (PHACE) experiment, we evaluated the responses of native mixed-grass prairie to 1 year of CO₂ enrichment (2006) (from present ambient (385 p.p.m.v.) to elevated (600 p.p.m.v.) CO₂, treatments c and C respectively), followed by 3 more years (2007–2009) of combined CO₂ enrichment and warming (present ambient, and elevated, temperatures (1.5/3.0 °C warmer canopy day/night), treatments t and T respectively) using free-air CO₂ enrichment (FACE) technology for CO₂ (ref. 17) and T-FACE technology for warming¹⁸ (Supplementary Appendix I). CO₂ enrichment and warming had opposing effects on soil water balance. Elevated CO₂ (Ct) increased annual SWC (17.3% averaged over 2007–2009), whereas warming reduced it (13.1%), such that there was no difference between control (ct; 15.5%) and CO₂-enriched and warmed plots (CT; 15.6%) (Supplementary Appendix II). The same pattern of compensating treatment effects on SWC was observed in the early growing season (Fig. 1), the time of year when SWC most controls annual productivity in this grassland¹⁹. These results are similar to but more consistent than those reported in a semi-arid Australian temperate grassland²⁰, and illustrate that the water conservation effects

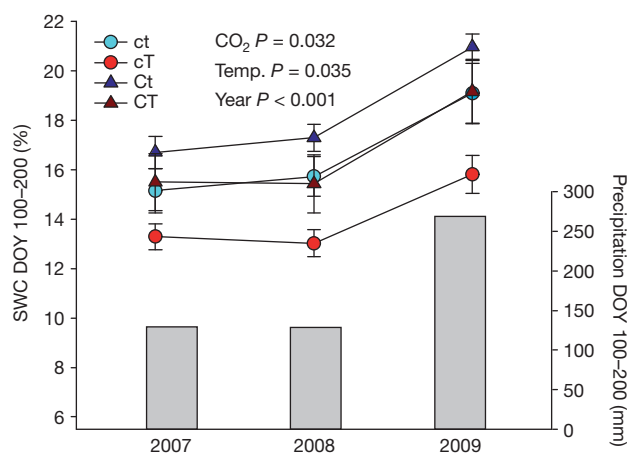


Figure 1 | Responses of SWC to CO₂ and warming. Average and s.e.m. (error bars) of volumetric SWC (5–25 cm depth) for plots exposed to present-day ambient CO₂ and temperature (ct), 1.5/3 °C day/night warming (cT), 600 p.p.m.v. CO₂ (Ct), and 600 p.p.m.v. CO₂ and 1.5/3 °C day/night warming (CT) (five replications per treatment). Data are averaged over days of year (DOY) 100–200, the early- to mid-growing season when soil water most limits productivity. Precipitation amounts for this same period are also presented. Significance ($P \leq 0.05$) for main effects and year are given in the figure.

¹USDA-ARS, Rangeland Resources Research Unit and Northern Plains Area, Fort Collins, Colorado 80526, USA. ²Department of Botany and Program in Ecology, University of Wyoming, Laramie, Wyoming 82071, USA. ³US Arid-Land Agricultural Research Center, USDA, Agricultural Research Service, Maricopa, Arizona 85238, USA. ⁴Departments of Botany, Renewable Resources, and Program in Ecology, University of Wyoming, Laramie, Wyoming 82071, USA. ⁵Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Sydney, New South Wales 2006, Australia.

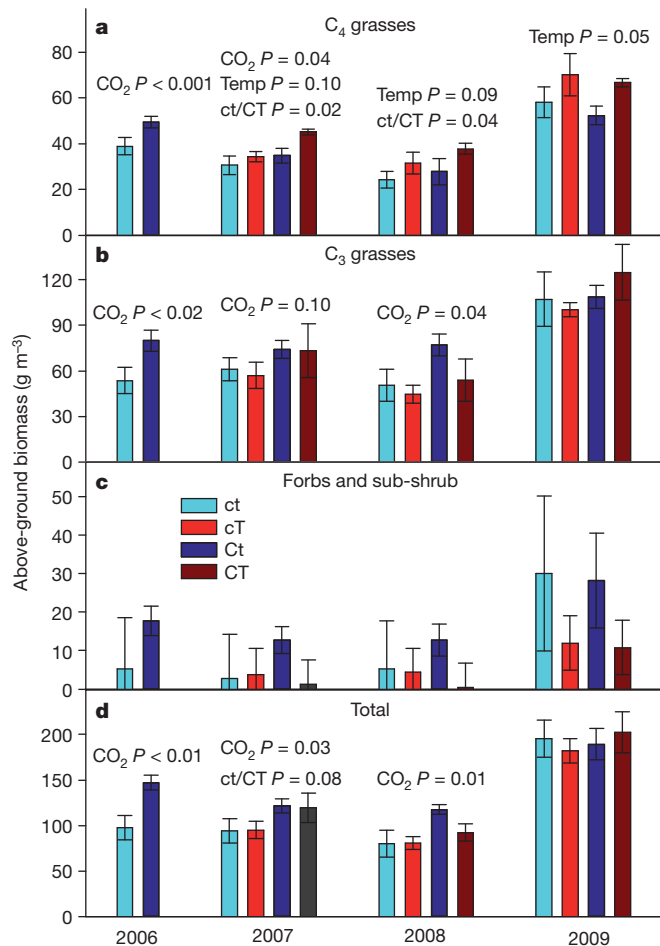
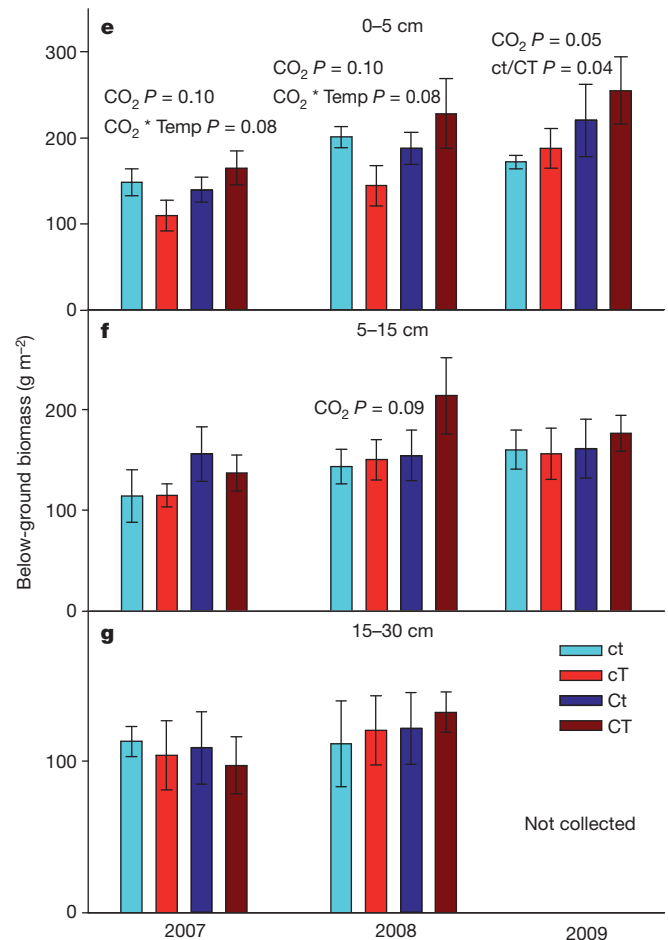


Figure 2 | Plant biomass responses to CO₂ and warming. Treatment effects (details in Fig. 1) on biomass in mid-July. AGB and s.e.m. (error bars) are given for C₄ (a) and C₃ grasses (b), forbs and sub-shrub (c) and total plant community (d). Below-ground biomass and s.e.m. (error bars) is given for soil depths of 0–5



(e), 5–15 (f) and 15–30 (g) cm. Results from 2006 include only the CO₂ treatment. Significant contrasts between present-day (ct) and future (CT) conditions are represented above histograms as ct/CT. P values are given for analyses of main effects, interactions and the ct/CT contrast when $P \leq 0.10$.

of elevated CO₂ can completely cancel the desiccating effects of moderately warmer temperatures.

Exposure of the prairie to 600 p.p.m.v. CO₂ increased peak total above-ground biomass (AGB, an estimate of above-ground net primary productivity) by an average 33% (Fig. 2d) in the first 3 years of the experiment when annual precipitation amounts were within 7% of the site 132-year average of 388 mm. We attribute this strong, positive response of AGB to CO₂ in this semi-arid grassland to the higher SWC in CO₂-enriched plots⁴. CO₂ enrichment had no effect on AGB in 2009, a relatively wet year for the site when annual precipitation was 17% higher than the long-term mean (Supplementary Appendix II). We suspect that higher SWC in 2009 (Fig. 1) minimized the potential water-relations benefit of CO₂ enrichment on plant productivity.

To evaluate further the effect of CO₂ on SWC and plant productivity, we examined how responses of AGB to increasing CO₂ varied with early-season soil matric potential (ψ_m) averaged during periods of active growth ($\psi_m \geq -1.5$ MPa). The ratio of AGB of plants exposed to 600 p.p.m.v. CO₂ to that of plants exposed to present-day CO₂ (biomass enhancement ratio) was used for this evaluation. We incorporated results from another CO₂ enrichment experiment in the Colorado shortgrass steppe to include the other semi-arid grassland of the western Great Plains²¹ (Supplementary Appendix I, Supplementary Fig. 1). A strong negative relationship was observed between ψ_m and biomass enhancement ratio (Fig. 3), which we attribute to (1) the enhancement of water use efficiency under elevated CO₂ and increasing water stress^{4,5}, and (2) the use of a physiologically relevant environmental driver (ψ_m rather than SWC; Supplementary Appendix I, soil water conversions)

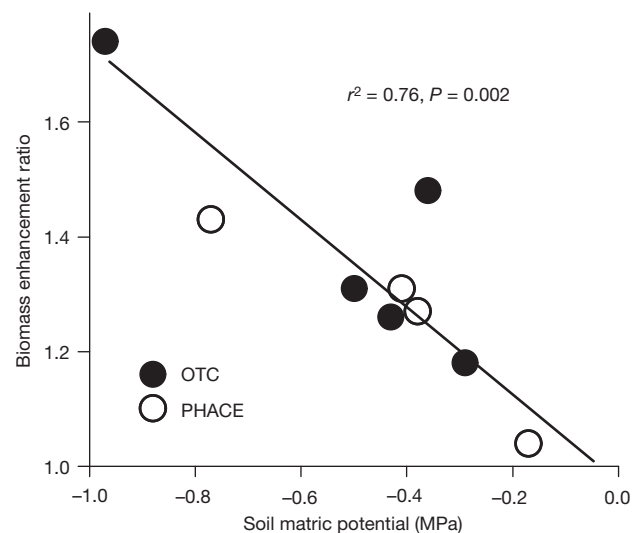


Figure 3 | Response of biomass enhancement ratio to soil matric potential. Effects of early-season (DOY 100–200) rooting zone soil matric potential (ψ_m) on biomass enhancement ratio, the ratio of mid-July harvested above-ground plant biomass in CO₂-enriched plots divided by plant biomass from ambient CO₂ plots. Data are from the PHACE experiment (4 years' data, $n = 4$), and from a previous open top chamber CO₂ enrichment experiment (5 years' data, $n = 5$) conducted on Colorado shortgrass steppe²¹. For further details, see Supplementary Appendix I, soil water conversions.

averaged over periods of physiological activity. This robust relationship from two different ecosystems suggests that CO₂ will increase plant productivity most when plants are water-limited yet still active.

The positive effect of CO₂ on production was also apparent in below-ground plant biomass. However, because root turnover in semi-arid perennial grasslands occurs every 5–7 years (ref. 22), we did not detect a significant ($P \leq 0.05$) effect on roots until the fourth year of our experiment (2009), when below-ground plant biomass in the 0–5 cm depth layer was 32% higher in CO₂-enriched plots (Fig. 2e).

Warming has the potential to increase production by extending the duration of the growing season, and by increasing nitrogen uptake, efficiency of nitrogen use²³ and biological activity. In our experiment, however, warming by itself did not significantly affect total AGB (Fig. 2d) or below-ground plant biomass (Fig. 2e–g), despite increased plant nitrogen content²⁴. Warming-induced desiccation in this semi-arid mixed-grass prairie (Fig. 1) and the preponderance of cool-season C₃ grasses (55% of AGB) probably limited its ability to respond positively to warming.

Although water was a dominant driver in the responses of both C₃ and C₄ grasses to CO₂ and temperature, there was also evidence of unique functional group responses. C₃ grass AGB was 34% greater in CO₂-enriched compared with ambient plots from 2006 to 2008 ($P < 0.05$ in 2006 and 2008; $P = 0.1$ in 2007) (Fig. 2b), but did not respond to warming. In contrast, AGB of C₄ grasses was 28% greater in CO₂-enriched plots in the first 2 years, but was also greater under warming ($P \leq 0.10$), most notably the wet year (2009) when C₄ AGB was 23% greater in heated plots ($P = 0.05$) (Fig. 2a). Comparing present (ct) to future (CT) conditions, C₃ grasses did not respond, but C₄ AGB significantly increased with the combination of warming and CO₂ enrichment in 2007 and 2008 (Fig. 2). These results are consistent with our physiological understanding of C₃ and C₄ plant functional types^{8,13}, but they demonstrate for the first time in a field experiment a potential advantage for C₄ grasses in a CO₂-enriched and warmer future. Further research will be needed to determine the extent to which these results can be applied to other C₃ and C₄ grasses and grassland ecosystems (Supplementary Appendix III, C₄ versus C₃ responses). Small population sizes and high year-to-year variability limited our capacity to detect treatment effects on forbs and subshrubs (Fig. 2c).

Our results suggest that even with modest increases in plant growth (2007), rising atmospheric CO₂ may counter the pervasive desiccating effects of warming in semi-arid grassland ecosystems. The most likely explanation is a CO₂-induced suppression of transpiration⁴. A modelling analysis using our site's weather data shows the effects of different hypothetical increases in canopy resistance to water loss (r_c) such as might occur from stomatal closure, and different temperature increases on grass reference evapotranspiration rate (ET_{ref}, Fig. 4), the evapotranspiration rate of non-stressed grassland. Differences in ET_{ref} reflect not only the rate of evapotranspiration when soil water is readily available to plants, but also how long grass can continue to transpire and grow after major rain events before exhausting its soil water supply. At ambient CO₂, a daytime warming of 1.5 °C (our cT treatment daytime temperature; roughly 90% of evapotranspiration occurs during daytime) is predicted to increase ET_{ref} by 7%. At ambient temperature, an increase in r_c of 30% would be required to decrease ET_{ref} by a similar amount, 6%. This condition is similar to our Ct treatment (Fig. 4); leaf gas exchange measurements on the dominant C₃ and C₄ grasses of this²⁵ and other²⁶ natural ecosystems suggest a 30–40% increase in stomatal resistance occurs in native grasses under CO₂-enriched and non-stressed conditions. Thus, when daytime warming of 1.5 °C is combined with an increase of 30% in r_c (our CT treatment), the temperature effect and the CO₂ effect through increased stomatal resistance almost exactly offset one another, so that an increase of about only 1% in ET_{ref} is predicted. This modelling result mirrors the observed offsetting effects of CO₂ and temperature on SWC (Figs 1 and Supplementary Figure 4). The analysis does not account for treatment

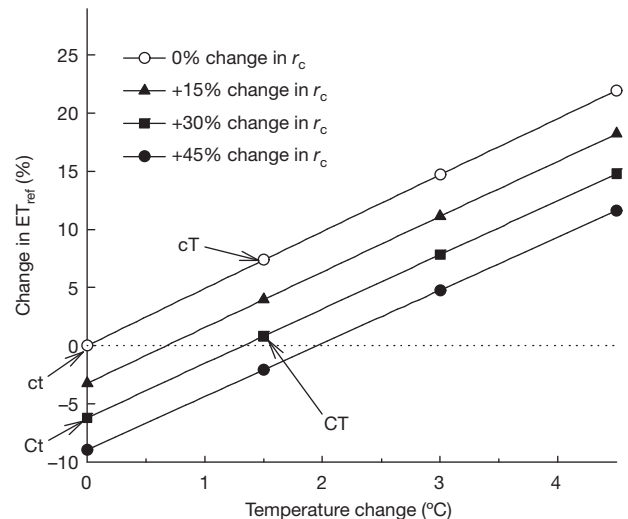


Figure 4 | Percentage changes in ET_{ref} for a grass surface as affected by temperature and changes in r_c . Percentage changes in ET_{ref} calculated using the American Society of Civil Engineers standardized evapotranspiration equation versus a range of temperature increases following Kimball²⁸. Calculations used observed weather data from the PHACE project for the 1 April – 16 October 2007, 2008 and 2009 growing seasons. The total calculated ET_{ref} for the three seasons was 2,490 mm (average of 5.1 mm d⁻¹). Calculations were done for zero, +15, +30 and +45% changes in canopy resistance to water vapour (r_c), as might be expected from CO₂-induced stomatal closure. For further details, see Supplementary Appendix III, Supplementary Fig. 5.

effects on leaf area, although such differences were probably small compared with 30–40% increases in stomatal resistance we would expect under CO₂ enrichment (Supplementary Appendix III, evapotranspiration modelling). A protracted drought that depletes soil water stores will of course eventually eliminate these CO₂-induced water savings; we have yet to experience such conditions in this experiment. Thus, although rising CO₂ may ameliorate the effects of warming-induced drought through plant stomatal closure, it is unlikely to eliminate severe and protracted droughts predicted for rangelands in regions like southwestern North America or the Mediterranean, where both substantive temperature increase and reduced precipitation are predicted^{1,2}.

The particular responses reported here are to a single set of simulated global change conditions that are within the prediction possibilities of the Intergovernmental Panel on Climate Change for the latter half of this century²⁷. The manipulations are not without their artefactual issues, which must be considered in scaling such information temporally and beyond the experimental site (Supplementary Appendix III, warming & CO₂ methodologies). Nevertheless, our results clearly illustrate the importance of compensating CO₂ and warming effects in semi-arid ecosystems. These contrasting, water-mediated effects must be accounted for in accurately predicting the susceptibility of such systems to climate change.

METHODS SUMMARY

Experiment. The experiment was conducted in northern mixed-grass prairie west of Cheyenne, Wyoming, USA. It had a factorial combination of two levels of CO₂ (385 p.p.m.v. and elevated 600 p.p.m.v.), and two temperature (present ambient, and elevated (1.5/3.0 °C warmer day/night)) regimes, with five replications each. Treatments were randomly assigned to 20 3.3-m diameter circular plots in two soil types (blocks). Details of the experimental site and set up are available elsewhere²⁴.

AGB was measured by species in 2005 (the year before CO₂ treatments), in 2006 (the first year of CO₂ treatments) and from 2007 to 2009 (the years of combined elevated CO₂ and warming treatments) during mid-July. A metal wire grid divided into 24–25 cm × 25 cm quadrats (1.5 m² total) was placed over each plot and vegetation in every other quadrat (12 in total) was clipped to the crown, dried at 60 °C and weighed. Plant species were subsequently grouped into three functional groups for analyses: C₄ grasses, C₃ grasses, and Forbs and sub-shrub.

Beginning in 2007, three soil cores (30 cm deep, 5 cm diameter) were collected at about the same time as the AGB samples from each plot, divided into 0–5, 5–15 and 15–30 cm depths, and the three core samples composited into one sample per depth. Fine roots (less than 1 mm) were handpicked from the composited samples, and root fragments were washed, weighed and ash-corrected.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 December 2010; accepted 7 June 2011.

Published online 3 August 2011.

- Wang, G. Agricultural drought in a future climate: results from 15 global change models participating in the IPCC 4th assessment. *Clim. Dyn.* **25**, 739–753 (2005).
- Seager, R. & Vecchi, G. A. Greenhouse warming and the 21st century hydroclimate of southwestern North America. *Proc. Natl Acad. Sci. USA* **107**, 21277–21282 (2010).
- Woodhouse, C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W. & Cook, E. R. A 1,200-year perspective of 21st century drought in southwestern North America. *Proc. Natl Acad. Sci. USA* **107**, 21283–21288 (2010).
- Morgan, J. A. *et al.* Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO₂. *Oecologia* **140**, 11–25 (2004).
- Leakey, A. D. B. Rising atmospheric carbon dioxide concentration and the future of C₄ crops for food and fuel. *Proc. R. Soc. Lond. B* **276**, 2333–2343 (2009).
- Frelich, L. E. & Reich, P. B. Will environmental changes reinforce the impact of global warming on the prairie–forest border of central North America? *Front. Ecol. Environ.* **8**, 371–378 (2010).
- Piao, S. *et al.* Changes in climate and land use have a larger direct impact than rising CO₂ on global river runoff trends. *Proc. Natl Acad. Sci. USA* **104**, 15242–15247 (2007).
- Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C-4 photosynthesis, atmospheric CO₂ and climate. *Oecologia* **112**, 285–299 (1997).
- Asner, G. P., Elmore, A. J., Olander, L. P., Martin, R. E. & Harris, A. T. Grazing systems, ecosystem responses, and global change. *Annu. Rev. Environ. Resour.* **29**, 261–299 (2004).
- Suttie, J. M., Reynolds, S. G. & Batello, C. *Grasslands of the World* (FAO, 2005).
- Noy-Meir, I. Desert ecosystems: environment and producers. *Annu. Rev. Ecol. Syst.* **4**, 25–51 (1973).
- McNaughton, K. G. & Jarvis, P. G. Effects of spatial scale on stomatal control of transpiration. *Agric. For. Meteorol.* **54**, 279–301 (1991).
- Epstein, H. E. *et al.* The relative abundance of three plant functional types in temperate grasslands and shrublands of North and South America: effects of projected climate change. *J. Biogeogr.* **29**, 875–888 (2002).
- Polley, H. W. Implications of rising atmospheric carbon dioxide concentration for rangelands. *J. Range Mgmt* **50**, 562–577 (1997).
- Semmartin, M., Aguiar, M. R., Distel, R. A., Moretto, A. S. & Ghersa, C. M. Litter quality and nutrient cycling affected by grazing-induced species replacements along a precipitation gradient. *Oikos* **107**, 148–160 (2004).
- Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K. & DeJong, D. D. NDVI, C₃ and C₄ production, and distributions in Great Plains grassland land cover classes. *Ecol. Appl.* **7**, 59–78 (1997).
- Miglietta, F. *et al.* Free-air CO₂ enrichment (FACE) of a poplar plantation: the POPFACE fumigation system. *New Phytol.* **150**, 465–476 (2001).
- Kimball, B. A. *et al.* Infrared heater arrays for warming ecosystem field plots. *Glob. Change Biol.* **14**, 309–320 (2008).
- Derner, J. D. & Hart, R. H. Grazing-induced modifications to peak standing crop in northern mixed-grass prairie. *Rangeland Ecol. Mgmt* **60**, 270–276 (2007).
- Hovenden, M. J. *et al.* Influence of warming on soil water potential controls seedling mortality in perennial but not annual species in a temperate grassland. *New Phytol.* **180**, 143–152 (2008).
- Morgan, J. A., Milchunas, D. G., LeCain, D. R., West, M. & Mosier, A. R. Carbon dioxide enrichment alters plant community structure and accelerates shrub growth in the shortgrass steppe. *Proc. Natl Acad. Sci. USA* **104**, 14724–14729 (2007).
- Milchunas, D. G., Morgan, J. A., Mosier, A. R. & LeCain, D. R. Root dynamics and demography in shortgrass steppe under elevated CO₂, and comments on minirhizotron methodology. *Glob. Change Biol.* **11**, 1837–1855 (2005).
- Luo, Y., Sherry, R., Zhou, X. & Wan, S. Terrestrial carbon-cycle feedback to climate warming: experimental evidence on plant regulation and impacts of biofuel feedstock harvest. *Glob. Change Biol.* **1**, 62–74 (2009).
- Dijkstra, F. A. *et al.* Contrasting effects of elevated CO₂ and warming on nitrogen cycling in a semiarid grassland. *New Phytol.* **187**, 426–437 (2010).
- LeCain, D. R., Morgan, J. A., Mosier, A. R. & Nelson, J. A. Soil and plant water relations determine photosynthetic responses of C₃ and C₄ grasses in a semi-arid ecosystem under elevated CO₂. *Ann. Bot. (Lond.)* **92**, 41–52 (2003).
- Ward, S. J. E., Midgley, G. F., Jones, M. H. & Curtis, P. S. Responses of wild C₄ and C₃ grass (Poaceae) species to elevated atmospheric CO₂ concentrations: a meta-analytic test of current theories and perceptions. *Glob. Change Biol.* **5**, 723–741 (1999).
- Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC* (eds Solomon, S. *et al.*) 747–845 (Cambridge Univ. Press, 2007).
- Kimball, B. A. in *Irrigation of Agricultural Crops (Agronomy Monograph No. 30)* 2nd edn (eds Lascano, R. J. & Sojka, R. E.) 627–653 (American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, 2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Smith for installation and operation of the PHACE experiment, E. Hardy for assistance in installation, A. Eden and C. Brooks for data collection and analysis, F. Miglietta for advice and help on installation of the FACE system, and R. Seager, A. Leakey, B. Cook and G. Wang for comments on the manuscript. The work was supported by the US Department of Agriculture–Agricultural Research Service Climate Change, Soils & Emissions Program, the US Department of Agriculture–Cooperative State Research, Education, and Extension Service Soil Processes Program (grant no. 2008-35107-18655), the US Department of Energy's Office of Science (Biological and Environmental Research) through the Western Regional Center of the National Institute for Climatic Change Research at Northern Arizona University, and the National Science Foundation (DEB no. 1021559). Mention of commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

Author Contributions J.A.M., E.P., D.M.B., B.A.K., D.G.W. and M.W. conceived the study. J.A.M., D.R.L., E.P., D.M.B., Y.C., D.G.W., J.H.-W. and F.A.D. performed the experiment. B.A.K. designed the warming system and conducted the evapotranspiration analysis. J.A.M. wrote the paper and the remaining authors contributed to the writing. Statistical analyses using SAS/STAT software, version 9.2, Proc GLIMMIX were performed by M.W. and J.A.M. The regression analyses using JMP software were performed by D.M.B. and J.A.M. Figures were developed by D.R.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.M. (jack.morgan@ars.usda.gov).

METHODS

Site description. The experiment was conducted at the US Department of Agriculture-Agricultural Research Service High Plains Grasslands Research Station, west of Cheyenne, Wyoming, USA (41° 11' N, 104° 54' W), elevation 1930 m. The ecosystem is a northern mixed-grass prairie, with a plant community of about 55% cool-season C₃ grasses (mostly *Pascopyrum smithii* and *Hesperostipa comata*, both perennials), 25% warm-season C₄ grasses (almost exclusively the perennial *Bouteloua gracilis*), and 20% sedges, forbs and small shrubs. Annual precipitation is 384 mm, mean air temperatures are 17.5 °C in summer and −2.5 °C in winter. The average annual wind speed is 6 m s^{−1}, with a growing season wind speed of 4.1 m s^{−1}.

A 2.4-ha site, which had been grazed by cattle at least since 1974, was fenced to prevent cattle entering in 2005. Soils are an Ascalon variant loam (fine-loamy, mixed-mesic) at the north end of the study site, and an Altvan loam (fine-loamy over sandy, mixed-mesic) at the south end.

Experimental layout. The experiment has a factorial combination of two levels of CO₂ (present ambient (385 p.p.m.v.), and elevated (600 p.p.m.v.), c and C respectively), and two temperature (present ambient, and elevated (1.5/3.0 °C warmer day/night), t and T respectively) regimes, with five replications making 20 experimental plots. Assuming an annual increase of CO₂ concentration of 2.3 p.p.m.v., our treatment combinations represent CO₂ and temperatures expected near the end of this century. We chose differential day/night temperatures based on the Fourth Intergovernmental Panel on Climate Change assessment, which suggests that daily minimal temperatures will increase faster than maximal temperatures²⁷. Furthermore, the equilibrium global mean surface air temperature for a doubling of atmospheric CO₂ is likely to lie between 2 and 4.5 °C. For our target CO₂ concentration of 600 p.p.m.v., our average day/night temperature of 2.25 °C lies at the upper third of a range we calculate to be between approximately 1.1–2.5 °C. The four treatment combinations were designated as follows: ct, ambient CO₂, non-heated; Ct, elevated CO₂, non-heated; cT, ambient CO₂, heated; CT, elevated CO₂, heated. Treatments were randomly assigned to 20 plots in northern and southern soil-type field blocks. Details are available elsewhere²⁴, including the FACE¹⁷ and infrared warming¹⁸ apparatus used to alter ambient CO₂ and temperature (see also Supplementary Appendix I).

Experimental rings and microclimate. To accommodate the FACE system, experimental plots were circular with a diameter of 3.3 m (area of 8.6 m²). Before instrumentation was installed, hexagonal trenches 60 cm deep were dug around the circumference of each plot and a plastic barrier installed to isolate treated plots hydraulically from outside non-treated soils. This hydraulic separation helped maintain any water-relations effects of the treatments.

Most of the plot area was maintained as native northern mixed prairie and was partitioned into soil and plant sampling areas, plus sections for other measurements, including SWC and air and soil temperature. Volumetric SWC was

measured daily using frequency domain reflectometry sensors at 10 and 20 cm soil depths (Sentek EnviroSMART sensors, Sentek Sensor Technologies) placed approximately 75 cm from the ring centres. Daily total precipitation was recorded with a rain gauge (Onset corp. S-RGA-M002).

Plant biomass. AGB was measured by species in 2005 (the year before CO₂ treatments), in 2006 (the first year of CO₂ treatments) and from 2007–2009 (the years of combined elevated-CO₂ and warming treatments) during mid-July. Mid-July is the approximate time of peak seasonal AGB, and provides a good estimate of above-ground net primary production. A metal wire grid divided into 24–25 cm × 25 cm quadrats (1.5 m² total) was placed over each plot (all sampled vegetation being within a metre of the ring centre), and vegetation in every other quadrat (12 in total) was clipped to the crown, dried at 60 °C and weighed. This defoliation protocol removed 50% of the green vegetation and represented the grazing removal for these grasslands. To correct for initial plot differences in species abundance, the 2005 pre-treatment harvest data were used to calculate overall means of each plant species' biomass across all 20 plots, plus the deviation for each species from that mean in every plot. These deviations were either added to or subtracted from individual species' biomass data collected in all plots for the subsequent treatment years, 2006–2009, thereby correcting for initial plot differences in plant species abundance. Plant species were subsequently grouped into three functional groups for analyses: C₄ grasses, C₃ grasses and sedges, and forbs and sub-shrub.

Beginning in 2007, near the time of the peak standing AGB harvest, three soil cores (30 cm deep, 5 cm diameter) were collected from each plot, divided into 0–5, 5–15 and 15–30 cm depths, and composited into one sample per depth. Fine roots (less than 1 mm) were handpicked from the composited samples, and root fragments were washed, weighed and ash corrected.

Statistical analyses. Data analysis used SAS/STAT software, version 9.2, Proc GLIMMIX, © 2002–2008 SAS Institute. Mean AGB and below-ground plant biomass were compared among all four treatments (five replications each) and years by fitting a repeated-measures general linear model to biomass. Plots within treatments represented subjects on which repeated measures were taken. Treatments and year were fitted as fixed-effects. Contrasts were constructed to estimate and test the effects of CO₂, warming and their interaction on biomass within years, owing to climatic differences among the four years. We hypothesized that increasing CO₂ would increase plant biomass (one-tailed test), but that warming would have no net effect on biomass (two-tailed test). Analyses of mean SWC (averaged annually for days of year 100–200) were conducted across years 2007–2009 when both CO₂ and warming treatments were operational. Regression curves of SWC and ψ_m with biomass enhancement ratio using data from the PHACE and open top chamber experiments were computed using JMP software, version 8.0.1, 2009 SAS Institute. Throughout, a *P* value of ≤ 0.05 is the standard for indicating significance, although marginal significance (*P* ≤ 0.10) is reported in figures, and sometimes commented on in the text as deemed appropriate.

Initial radiation of jaws demonstrated stability despite faunal and environmental change

Philip S. L. Anderson¹, Matt Friedman², Martin D. Brazeau^{3†} & Emily J. Rayfield¹

More than 99 per cent of the roughly 58,000 living vertebrate species have jaws¹. This major clade, whose members are collectively known as gnathostomes ('jawed mouths'), made its earliest definitive appearance in the Silurian period, 444–416 million years (Myr) ago, with both the origin of the modern (crown-group) radiation and the presumptive invasion of land occurring by the end of the Devonian period² (359 Myr ago). These events coincided with a major faunal shift that remains apparent today: the transition from Silurian ecosystems dominated by jawless fishes (agnathans) to younger assemblages composed almost exclusively of gnathostomes^{2,3}. This pattern has inspired several qualitative descriptions of the trophic radiation and ecological ascendance of the earliest jawed vertebrates^{3–7}. Here we present a quantitative analysis of functional variation in early gnathostome mandibular elements, placing constraints on our understanding of evolutionary patterns during this critical interval. We document an initial increase in functional disparity in the Silurian that stabilized by the first stage of the Devonian, before the occurrence of an Emsian (~400 Myr ago) oxygenation event implicated in the trophic radiation of vertebrates⁸. Subsequent taxonomic diversification during the Devonian did not result in increased functional variation; instead, new taxa revisited and elaborated on established mandibular designs. Devonian functional space is dominated by lobe-finned fishes and 'placoderms'; high disparity within the latter implies considerable trophic innovation among jaw-bearing stem gnathostomes. By contrast, the major groups of living vertebrates—ray-finned fishes and tetrapods—show surprisingly conservative mandibular morphologies with little indication of functional diversification or innovation. Devonian gnathostomes reached a point where they ceased to accrue further mandibular functional disparity before becoming taxonomic dominants relative to 'ostracoderm'-grade jawless fishes, providing a new perspective on classic adaptive hypotheses concerning this fundamental shift in vertebrate biodiversity.

Morphological disparity has often been used as a proxy for ecological/functional variety in the fossil record⁹. However, divergent anatomies can be biomechanically similar^{10–12}, leading to decoupling of morphological and functional metrics^{12–14}. To evaluate potential functional diversity, we have devised a measure of variation derived from biomechanical traits. We focus on the mandible because it is well characterized functionally in modern fishes¹⁵ and because it is the key character proposed by adaptive schemes to separate the gnathostomes from their jawless predecessors^{4,6}. We document how functional variation accumulates after the origin of jaws and test the hypothesis that, during their rise to dominance, gnathostomes outcompeted or displaced jawless fishes.

We assembled a data set consisting of 198 Silurian/Devonian gnathostome genera scored for 31 mandibular traits (11 continuous and 20 discrete), many of which have been shown to be functionally relevant in living taxa. These data were analysed using non-metric multidimensional scaling, yielding a multivariate 'function space'. Functional disparity was calculated across eight time bins (the late

Silurian plus the seven stages of the Devonian) using a variety of measures (sum of variances is presented here because it is relatively robust to variation in sample size¹⁶). We tested for differences between successive bins using multiple approaches (likelihood ratios and *t*-tests), and used a jackknifing procedure that pruned exceptional fossil assemblages from our data set to examine the sensitivity of results to 'Lagerstätten effects', biases arising from localities characterized by exceptional preservation or fossil abundance. Full details of our methodology are given in Supplementary Information.

Our results show remarkably stable levels of lower-jaw functional disparity during the Devonian (Fig. 1a). There is a trend of increasing disparity from the late Silurian to the end of the Early Devonian (the Emsian stage); however, by the earliest Devonian (the Lockhovian stage, ~415 Myr ago) functional mandibular disparity attained a level statistically indistinguishable from that found in the Late Devonian. When *P* values are adjusted for multiple comparisons, the only significant disparity shift between successive Devonian stages (corrected *P* value, 0.013) is a decline on entering the Middle Devonian (the Eifelian stage, ~398 Myr ago; Fig. 1a). This seems to reflect a reduction in the number of marine sites sampled in the Eifelian, combined with the Emsian Taemas/Wee Jasper fauna, which contains a series of extensively studied durophagous lungfishes with exceptionally robust jaws. When this site is removed in jackknife analyses, the drop in disparity is no longer significant (corrected *P* value, 0.112). When differences are assessed using likelihood ratios, none exceeds a threshold value of 8 (ref. 17; likelihood ratio range, 1.023–4.34). This pattern of functional stability arises despite an increase both in the number of mandibles sampled per bin over our study interval and in the taxonomic diversity of gnathostomes during the Devonian¹⁸, and is robust to the removal of exceptional Lagerstätten. Details of all analyses are given in Supplementary Information.

A breakdown of morphospace by stage shows that initial functional disparity in the late Silurian was localized in the lower-left and upper-right quadrants (Fig. 2). The former is populated by gnathostomes (for example stem sarcopterygians and ischnacanthid 'acanthodians') with gracile mandibles bearing slender, pointed cusps and characterized by low closing mechanical advantages, whereas the latter contains taxa (for example mesacanthid acanthodians) with toothless jaws. The Lockhovian expansion into the lower-right quadrant reflects the appearance of taxa (for example the lungfish *Diabolepis*) with robust jaws characterized by both high closing mechanical advantage and high second moment of area, two features consistent with durophagy. Subsequent mandibular evolution during the Devonian resulted in novel designs that elaborated existing ones, but did not substantially exceed the boundaries established earlier in the interval (Fig. 2). Specific anatomical correlates of our ordination axes are provided in Supplementary Table 2.

We examined patterns of partial disparity¹⁹ to dissect the contributions to overall functional variation made by major taxonomic assemblages: 'Acanthodii', an extinct assemblage sometimes called spiny sharks; Actinopterygii, or ray-finned fishes, related to most fishes alive today; Chondrichthyes, or cartilaginous fishes, the precursors of rays, sharks and ratfishes; 'Placodermi', armoured stem gnathostomes; and

¹Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. ²Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK. ³Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University in Berlin, 10115 Berlin, Germany. [†]Present address: NCB Naturalis, 2300 RA Leiden, The Netherlands.

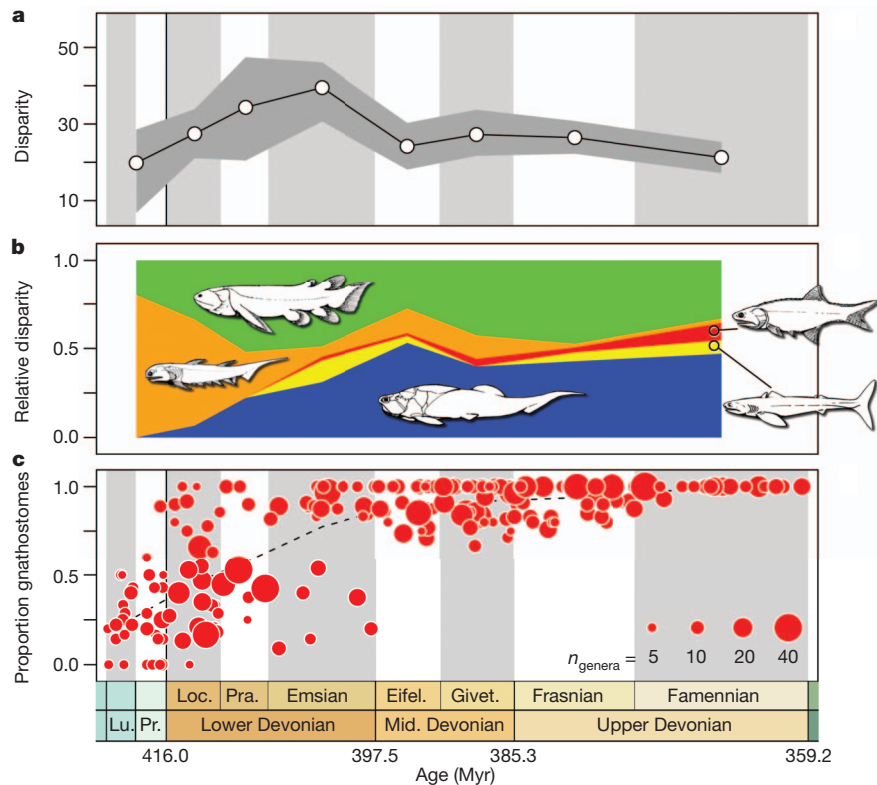


Figure 1 | Functional mandibular disparity among Silurian/Devonian gnathostomes. All horizontal axes show time as indicated at bottom.

a, Disparity (sum of variances) across eight time bins. The dark-grey region spans the 95% confidence intervals based on 1,000 bootstrap pseudoreplicates. **b**, Relative contributions (partial disparity) of major gnathostome groups to overall functional disparity. Orange, 'Acanthodii'; green, Sarcopterygii; blue, 'Placodermi'; yellow, Chondrichthyes; red, Actinopterygii. **c**, Faunal

composition data for the late Silurian and Devonian. Discs represent individual vertebrate assemblages plotted as a function of time and proportion of gnathostomes that comprise those faunas (discs jittered within time bins for clarity). The area of each disc is proportional to the total number of vertebrate genera represented, n_{genera} . Eifel., Eifelian stage; Givet., Givetian stage; Loc., Lockhovian stage; Lu., Ludlow series; Pr., Pridoli series; Pra., Pragian stage.

Sarcopterygii, or lobe-finned fishes, including lungfishes and terrestrial vertebrates. Sarcopterygians make a major and relatively stable contribution to mandibular disparity throughout our study interval (Fig. 1b). The earliest definitive digit-bearing tetrapods (*Acanthostega* and *Ichthyostega*) and their immediate relatives (*Ventastega*, *Elginerpeton*, *Densignathus* and *Metaxygnathus*) show a conservative pattern in mandibular function relative to sarcopterygian 'fishes' (Fig. 2). This is in contrast to studies that report a distinct shift in cranial suture patterns between aquatic taxa and early digit-bearing tetrapods²⁰, suggesting that different parts of the feeding system evolved in separate phases. By contrast, lungfishes are a major source of mandibular innovation among sarcopterygians, colonizing a new region of functional space in the Lockhovian and occupying it exclusively throughout the Devonian (Fig. 2).

'Placoderms' occupy 50% of realized Devonian gnathostome morphospace volume (Fig. 2), indicating that considerable functional differentiation occurred on the gnathostome stem, independently of the origin of modern jawed vertebrates. 'Placoderm' functional variety remains undiminished until their sudden extinction at the Devonian/Carboniferous boundary. A direct reading of our results implies that 'placoderms' were a major contributor to gnathostome disparity only from the Middle Devonian onward. We argue that this is an artefact; many 'placoderms' are known from the Early Devonian and, increasingly, the Silurian^{2,21}, but few preserve mandibular material that can be included in our analysis. This suggests that disparity measures for the Silurian and Early Devonian are probably conservative underestimates, and that the addition of more 'placoderms' of this age would only reinforce the pattern of early functional morphospace saturation.

The remaining three gnathostome assemblages—'acanthodians', chondrichthyans and actinopterygians—make relatively minor

contributions to overall disparity (Fig. 1b). 'Acanthodians' make a major contribution to functional variety during the Silurian, but their relative importance wanes substantially throughout the Devonian, reflecting a decline in functional diversity and apparent richness combined with the taxonomic ascendance of other gnathostome clades²². Actinopterygians, the most speciose group of living jawed fishes¹, contribute little to overall disparity in the Devonian, with their few representatives conservatively clustering within tightly delimited regions of function space (Fig. 2). This is consistent with previous arguments, drawn from taxonomic data, that actinopterygians were not ecologically diverse until the Carboniferous period²². The reliability of low measures of chondrichthyan partial disparity is unclear, however, because the numerous Devonian chondrichthyans known exclusively from dental remains²³ cannot be included in our analysis.

Our results provide a new perspective on debates concerning the Devonian shift from agnathan- to gnathostome-dominated fossil assemblages, a transition that remains apparent in modern vertebrate fauna^{1,22}. A range of schemes, largely derived from anecdotal evidence, have sprung up in response to this pattern of turnover^{3–7}. Faunal data clearly show that gnathostomes shared habitat space evenly with 'ostracoderm'-grade agnathans well into the Early Devonian (Fig. 1c). It is only during and after the Emsian that gnathostomes became taxonomic dominants in most fossil assemblages, several million years after they explored the extremes of realized Devonian mandibular function space. The offset between the trophic radiation of gnathostomes (Fig. 1a) and the taxonomic collapse of 'ostracoderms' (Fig. 1c), combined with the long history of coexistence between these two groups, argues against the direct ecological replacement of jawless fishes by jawed forms⁶. Furthermore, we find no evidence for a significant positive relationship between mandibular disparity within assemblages and the taxonomic

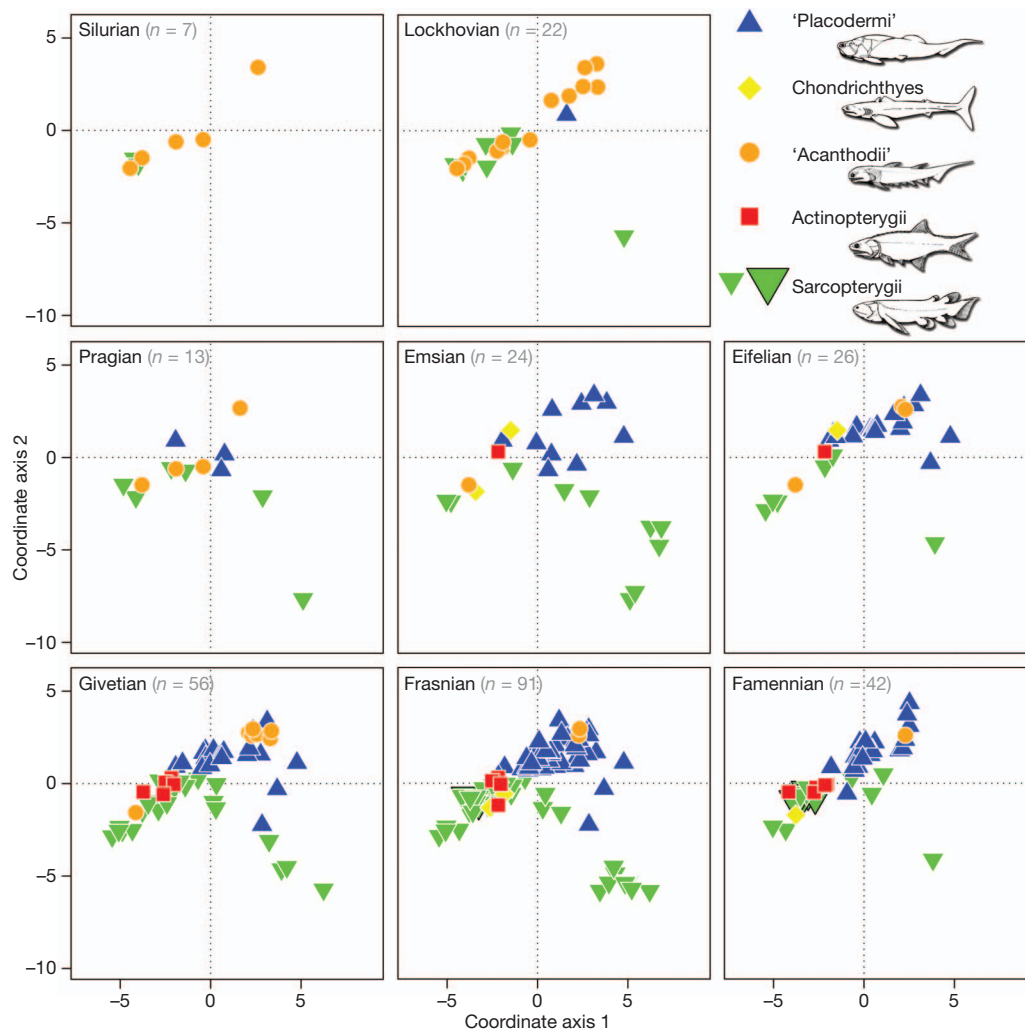


Figure 2 | Patterns of functional morphospace occupation for jawed vertebrates during the late Silurian and Devonian. Points represent averages for genera plotted on coordinate axes 1 and 2 of the non-metric multidimensional scaling ordination. Orange circles, 'Acanthodii'; green down-triangles, Sarcopterygii (tetrapods indicated by larger symbols outlined in

black); blue up-triangles, 'Placodermi'; yellow diamonds, Chondrichthyes; red squares, Actinopterygii. The area in the lower-right corner of the functional morphospace is occupied solely by lungfishes. See text and Supplementary Information for further details.

dominance of jawed vertebrates within those faunas (measured as the proportion of gnathostomes; Pearson and Kendall rank-order correlations, $P = 0.90\text{--}0.94$; weighted least-squares regression, $P = 0.40\text{--}0.44$). This pattern is inconsistent with the assertion⁶ that gnathostome diversification primarily reflects the ecological displacement of agnathans and subsequent refilling of their functional roles. Indeed, few of the feeding strategies apparently taken up by early gnathostomes, such as macropredation and durophagy, have any clear analogue among jawless taxa³. Instead, the proliferation of gnathostomes probably reflects new ecological opportunities associated with the origin of jaws and, perhaps just as importantly, the shift from benthic lifestyles to pelagic ones²⁴.

The Devonian radiation of gnathostomes is marked by an early plateau of functional disparity and a consistent range of mandibular variation in the face of major environmental perturbations, including fluctuating global temperatures²⁵, shifting continental weathering patterns²⁶, sea level changes²⁶ and a major extinction event²⁷. Various proposals have linked patterns of early vertebrate evolution to changes in atmospheric composition, with parallel claims that gnathostome diversification was driven by increases in oxygen levels during the Silurian²⁸ or the Devonian⁸. Silurian data are too scant to comment on the first proposal, but we note that gnathostomes achieved stable levels of functional diversity before the occurrence of an Emsian oxygenation event (~400 Myr ago) implicated in the trophic radiation

of vertebrates⁸. There is no significant shift in functional disparity between the Frasnian and Famennian stages (~375 Myr ago; Fig. 1) correlating with the Kellwasser event, which is classically identified as one of the 'Big Five' Phanerozoic extinctions²⁷. This event resulted in a global reef collapse²⁹; however, jawed vertebrates seem to be unaffected in terms of diversity²², functional disparity and potentially ecological variation. More broadly, our results imply constraints on the functional diversity of the gnathostome mandibular system, with early functional saturation and subsequent refilling of established biomechanical roles by new taxonomic actors.

Received 16 March; accepted 18 May 2011.

Published online 6 July 2011.

- Nelson, J. S. *Fishes of the World* 4th edn (Wiley, 2006).
- Janvier, P. *Early Vertebrates* (Clarendon, 1996).
- Janvier, P. & Bleick, A. in *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (ed. Long, J. A.) 67–86 (Johns Hopkins Univ. Press, 1993).
- Halstead, L. B. *The Pattern of Vertebrate Evolution* 26 (Oliver & Boyd, 1969).
- Colbert, E. H. *Evolution of the Vertebrates* 28 (Wiley, 1991).
- Carroll, R. L. *Vertebrate Paleontology and Evolution* 44 (Freeman, 1988).
- Pough, F. H., Janis, C. M. & Heiser, J. B. *Vertebrate Life* 8th edn (Pearson Benjamin Cummings, 2009).
- Dahl, T. W. *et al.* Devonian rise in atmospheric oxygen correlated to the radiations of terrestrial plants and large predatory fish. *Proc. Natl Acad. Sci. USA* **107**, 17911–17915 (2010).
- Foote, M. Morphological disparity in Ordovician–Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**, 320–344 (1994).

10. Alfaro, M. E., Bolnick, D. I. & Wainwright, P. C. Evolutionary dynamics of complex biomechanical systems: an example using the four-bar mechanism. *Evolution* **58**, 495–503 (2004).
11. Alfaro, M. E., Bolnick, D. I. & Wainwright, P. C. Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes. *Am. Nat.* **165**, E140–E154 (2005).
12. Wainwright, P. C. Functional versus morphological diversity in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **38**, 381–401 (2007).
13. Hulsey, C. D. & Wainwright, P. C. Projecting mechanics into morphospace: disparity in the feeding system of labrid fishes. *Proc. R. Soc. Lond. B* **269**, 317–326 (2002).
14. Anderson, P. S. L. Biomechanics, functional patterns, and disparity in Late Devonian arthrodires. *Paleobiology* **35**, 321–342 (2009).
15. Westneat, M. W. Evolution of levers and linkages in the feeding mechanisms of fishes. *Integr. Comp. Biol.* **44**, 378–389 (2004).
16. Ciampaglio, C. N., Kemp, M. & McShea, D. W. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* **27**, 695–715 (2001).
17. Royall, R. M. *Statistical Evidence: A Likelihood Paradigm* (Chapman & Hall, 1997).
18. Benton, M. J. *The Fossil Record 2* (Chapman & Hall, 1993).
19. Foote, M. Contributions of individual taxa to overall morphological disparity. *Paleobiology* **19**, 403–419 (1993).
20. Markey, M. J. & Marshall, C. R. Terrestrial-style feeding in a very early aquatic tetrapod is supported by evidence from experimental analysis of suture morphology. *Proc. Natl Acad. Sci. USA* **104**, 7134–7138 (2007).
21. Zhao, W.-J. & Zhu, M. Siluro-Devonian vertebrate biostratigraphy and biogeography of China. *Palaeoworld* **19**, 4–26 (2010).
22. Sallan, L. C. & Coates, M. I. End-Devonian extinction and a bottleneck in the early evolution of modern jawed vertebrates. *Proc. Natl Acad. Sci. USA* **99**, 8139–8144 (2002).
23. Ginter, M., Hampe, O. & Duffin, C. J. *Handbook of Paleoichthyology Vol. 3D: Chondrichthyes Paleozoic Elasmobranchii* (Pfeil, 2010).
24. Klug, C. *et al.* The Devonian nekton revolution. *Lethaia* **43**, 465–477 (2010).
25. Joachimski, M. M. *et al.* Devonian climate and reef evolution: insights from oxygen isotopes in apatite. *Earth Planet. Sci. Lett.* **284**, 599–609 (2009).
26. van Geldern, R. *et al.* Carbon, oxygen and strontium isotope records of Devonian brachiopod shell calcite. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **240**, 47–67 (2006).
27. Raup, D. M. & Sepkoski, J. J. Jr. Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503 (1982).
28. Qu, Q. M., Zhu, M. & Zhao, W. J. Silurian atmospheric O₂ changes and the early radiation of gnathostomes. *Palaeoworld* **19**, 146–159 (2010).
29. Copper, P. Reef development at the Frasnian/Famennian mass extinction boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **181**, 27–65 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Donoghue, M. Rücklin and M. Ruta for discussions. This work was supported by Royal Society and Marie-Curie Actions fellowships, awarded to P.S.L.A.; a Fell Fund award to M.F.; NERC grant NE/G016623/1, awarded in part to E.J.R.; and a QRNT postdoctoral fellowship, to M.D.B.

Author Contributions P.S.L.A. designed and led the study, developed the biomechanical traits, collected data, performed the disparity and multivariate analyses, and wrote the paper. M.F. gathered data, wrote analytical code, performed the faunal analyses, drafted figures and wrote the paper. M.D.B. gathered data and produced illustrations. E.J.R. gave advice on biomechanical theory. All authors contributed to interpretation of the results and edited the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S.L.A. (phil.anderson@bristol.ac.uk).

Neuronal basis of age-related working memory decline

Min Wang¹, Nao J. Gamo¹, Yang Yang¹, Lu E. Jin¹, Xiao-Jing Wang¹, Mark Laubach^{1,2}, James A. Mazer¹, Daeyeol Lee¹ & Amy F. T. Arnsten¹

Many of the cognitive deficits of normal ageing (forgetfulness, distractibility, inflexibility and impaired executive functions) involve prefrontal cortex (PFC) dysfunction^{1–4}. The PFC guides behaviour and thought using working memory⁵, which are essential functions in the information age. Many PFC neurons hold information in working memory through excitatory networks that can maintain persistent neuronal firing in the absence of external stimulation⁶. This fragile process is highly dependent on the neurochemical environment⁷. For example, elevated cyclic-AMP signalling reduces persistent firing by opening HCN and KCNQ potassium channels^{8,9}. It is not known if molecular changes associated with normal ageing alter the physiological properties of PFC neurons during working memory, as there have been no *in vivo* recordings, to our knowledge, from PFC neurons of aged monkeys. Here we characterize the first recordings of this kind, revealing a marked loss of PFC persistent firing with advancing age that can be rescued by restoring an optimal neurochemical environment. Recordings showed an age-related decline in the firing rate of DELAY neurons, whereas the firing of CUE neurons remained unchanged with age. The memory-related firing of aged DELAY neurons was partially restored to more youthful levels by inhibiting cAMP signalling, or by blocking HCN or KCNQ channels. These findings reveal the cellular basis of age-related cognitive decline in dorsolateral PFC, and demonstrate that physiological integrity can be rescued by addressing the molecular needs of PFC circuits.

Our society is rapidly ageing, with the number of seniors in the United States expected to double by 2050 (United States census, <http://www.census.gov/population/www/pop-profile/elderpop.html>). At the same time, the information age requires increasing organizational skills to deal with even basic needs such as medical care and paying bills. However, executive and working memory functions decline early in the normal ageing process^{10–13}, beginning in middle age^{14,15}. Thus, cognitive changes with advancing age may be costly, forcing retirement from demanding careers and jeopardizing the ability to live independently in an increasingly complex society. Ageing monkeys provide an ideal model to reveal the neurobiology of normal ageing, as they have a highly developed PFC, but are not subject to age-related dementias¹⁶. Thus, one can be certain that cognitive changes are the result of normal ageing and not incipient Alzheimer's disease. Like humans, monkeys begin to develop deficits in executive function as early as middle age¹⁷. Both aged monkeys^{18,19} and humans^{3,20} are impaired on working memory tasks that require constant updating of the contents of memory (Supplementary Information), bringing to mind information from longer-term stores (for example, where did I leave my car keys this time?), or keeping in mind a recent event (for example, remembering a new phone number).

In primates, spatial working memory depends on the highly evolved dorsolateral PFC⁶ (Fig. 1a). Spatial working memory performance (Fig. 1b) relies on networks of pyramidal neurons that interconnect at dendritic spines (Fig. 1c), and excite each other to keep information 'in mind', that is, generating persistent spiking activity over a delay

period in a working memory task⁶ (Fig. 1d). This ability to maintain information that is no longer in the environment is a fundamental process needed for abstract thought and flexible responding⁶. Intracellular signalling pathways modulate the physiological strength of these recurrent, excitatory PFC network connections⁹. Recent data show that increased cAMP signalling weakens network connectivity by opening potassium channels, whereas inhibiting cAMP signalling and/or closing these channels strengthens connectivity and cognitive ability⁹ (Fig. 1e). Specifically, cAMP signalling seems to weaken persistent firing and impair working memory by increasing the open state of HCN (hyperpolarization-activated cyclic nucleotide-gated) channels that are localized on spines where networks interconnect⁸. Recent data indicate that HCN channels may also gate synaptic inputs through interactions with KCNQ channels, whose open state is increased by cAMP-activating protein kinase A (PKA)²¹. Studies indicate that cAMP signalling is disinhibited in the aged PFC²². Noradrenergic α_{2A} receptor inhibition of cAMP may be reduced from loss of α_{2A} receptors in the aged PFC²³, and decreased excitation of noradrenergic neurons²⁴.

There have been few electrophysiological recordings from aged PFC neurons owing to the demanding nature of this procedure. Recordings from rat orbital PFC found reduced flexibility in aged neurons²⁵. However, there have been no *in vivo* recordings from the aged dorsolateral PFC, even though behavioural data indicate that this region is particularly vulnerable to normal ageing. *In vitro* recordings from dorsolateral PFC neurons found relatively subtle changes in excitability with advancing age²⁶, but their consequences for executive function must be observed in a cognitively engaged circuit. Here we perform the first physiological characterization of PFC neuronal response during a working memory task in young adult, middle-aged and aged monkeys.

Monkeys (*Macaca mulatta*, $n = 6$) were trained to perform a spatial working memory task in which they have to remember a spatial location over a brief delay period; the spatial location changes randomly on each trial (Fig. 1a). Two animals were young adults (7- and 9-year-old males), two were middle-aged (12- and 13-year-old males), and two were aged (17-year-old male, 21-year-old female). Short delays (2.5 s) were used in all age groups to ensure similar performances (>85% per cent correct) across age groups. Neurons ($n = 301$) were recorded from area 46, the dorsolateral PFC subregion most needed for visuospatial working memory (Fig. 1a). Neurons were characterized based on task-related firing as responsive during (1) the visuospatial cue period, (2) the delay period when the spatial position was being remembered, and/or (3) the motor response period. Some neurons fired only during cue presentation (CUE cells, $n = 28$), whereas most neurons fired during the delay period as well as to the cue and/or response periods (DELAY cells, $n = 273$). Persistent firing during the delay period is of particular interest, as it is required for working memory⁶. Many PFC DELAY neurons increased their activity during the memory of one spatial location (its preferred direction), but not other locations (the 'anti-preferred' direction, 180° away from the preferred direction; Fig. 1d).

¹Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ²The John B. Pierce Laboratory, New Haven, Connecticut 06510, USA.

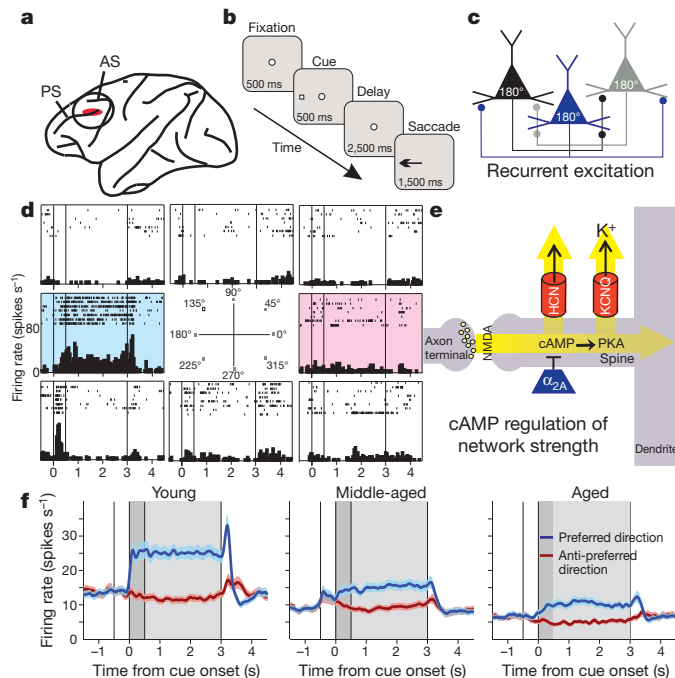


Figure 1 | Age-related changes in the PFC networks that subserve working memory. **a**, The region of the dorsolateral PFC most needed for spatial working memory, and the site of recordings. AS, arcuate sulcus; PS, principal sulcus. **b**, The oculomotor delayed response (ODR) spatial working memory task. The monkey fixates on the central stimulus and maintains fixation for the duration of the trial. A cue is briefly presented in one of eight locations, followed by a delay period (2.5 s) in which no spatial information is present. At the end of the delay period, the fixation spot disappears, and the monkey makes an eye movement (saccade) to the remembered location for a juice reward. The cue position randomly changes on subsequent trials. **c**, A diagram of the recurrent excitatory networks subserving working memory. Pyramidal cells with similar spatial tuning excite each other to maintain persistent firing across the delay period⁶. These networks are concentrated in deep layer III⁶. Spatial tuning is enhanced by GABAergic lateral inhibition (not shown). **d**, An example of a dorsolateral PFC DELAY neuron with spatially tuned, persistent firing during the delay period. This neuron shows increased firing for the cue, delay and response for the neuron's preferred direction (highlighted in blue), but not for nonpreferred directions (white backgrounds). The anti-preferred direction opposite to the neuron's preferred direction is shown in red; note that subsequent figures show only the preferred and anti-preferred directions for the sake of brevity. **e**, Pyramidal cells synapse on spines where cAMP–PKA signalling regulates the open state of HCN and KcNQ channels, and thus modulates the strength of network connections⁹. **f**, Population average activity for the dorsolateral PFC DELAY neurons recorded in each age group (102, 101 and 70 neurons for young, middle-aged and old monkeys, respectively). Colours indicate the activity during the trials in which the cue was presented in the neuron's preferred (blue) and anti-preferred (red) directions; the darker grey background refers to the cue period; the lighter grey background to the delay period. Error envelope represents s.e.m.

The firing of DELAY cells was markedly reduced with advancing age (Figs 1f and 2). Figure 1f portrays the differences in firing rates across the population of DELAY neurons in young, middle-aged and aged animals (for individual examples of DELAY neurons in young, middle-aged and aged monkeys, see Supplementary Fig. 1). There was a significant decline in the spontaneous firing rate of DELAY cells, as well as a marked decline in task-related firing. Figure 2 shows a steep decline in the firing rates of DELAY cells across the age span (t -test on age variable in regression analysis, $P < 10^{-4}$ for all epochs), with older animals showing a restricted range of lower firing rates (Supplementary Fig. 2). This age-related activity decline persisted throughout the 2.5-s delay period (without main effect of epoch (0.5 s) or age \times epoch interaction in repeated measures ANOVA, $P > 0.25$). Additional control analyses showed that age-related decline in the firing rate of

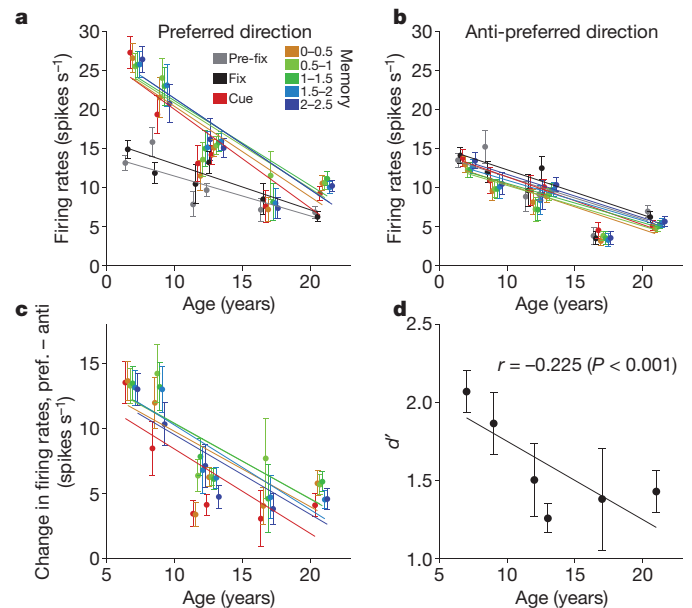


Figure 2 | Age-dependent decline in the spatially tuned, persistent firing of dorsolateral PFC DELAY neurons. **a**, Marked reduction of dorsolateral PFC DELAY activity for the neurons' preferred direction with advancing age. Activity of individual neurons of each animal was averaged separately for the last 0.5 s during the intertrial interval (pre-fix, grey), the fixation period (fix, black), the cue period (cue, red), and the delay period for the neuron's preferred direction. Firing during the delay period is represented in a successive series of 0.5 s intervals (colour coded from yellow to blue). Lines were obtained using linear regression. **b**, Firing rates during the delay period for the anti-preferred direction of the same neurons shown in **a**. There was a significant age-related decline in all epochs, but it was less prominent than the decline in firing for the preferred direction during the delay period. Colour coding for each 0.5 s interval is as in **a**. **c**, Age-related decline in spatial tuning, whereby the difference between firing for the preferred versus anti-preferred directions during the delay period declines with advancing age. Colour coding is as in **a**. **d**, Age-related decline in d' , that is, the ability to distinguish preferred from anti-preferred spatial directions based on firing rate patterns during the entire delay period. Error bars represent s.e.m.

DELAY cells is not due to a sampling bias during the recording experiment (see Supplementary Information). The age-related decline in firing rate was particularly prominent during the cue and delay periods for the neuron's preferred direction (Fig. 2a); the decline in firing for the anti-preferred direction (Fig. 2b) or before target onset (Fig. 2a) was less pronounced. Consequently, the difference in delay-related firing for the neuron's preferred direction versus its anti-preferred direction eroded with increasing age (t -test, $P < 10^{-5}$, for cue period and every 0.5 s epoch in the delay period; Fig. 2c), largely due to reduced firing for the neuron's preferred direction (Fig. 2a). This led to a reduction in d' with advancing age (t -test on age versus d' correlation coefficient, $P < 0.0001$), that is, a reduced ability to distinguish the preferred from anti-preferred directions during the delay period when spatial information was held in working memory (Fig. 2d). These results are consistent with studies showing impairment in spatial working memory in aged monkeys at relatively short (for example, 5 s) delays¹⁸, and single-unit data as well as neural circuit modelling indicate that inadequate PFC recurrent network firing underlies the deficits in PFC cognitive function observed in ageing monkeys and humans (Supplementary Figs 3, 4 and Supplementary Information).

In contrast to DELAY neurons, which showed prominent decline in firing with advancing age, there were no age-related changes in the firing rates of PFC CUE cells that responded specifically to the spatial cue (Fig. 3). The average firing rate of these neurons for the preferred direction during the cue period was 26.7 ± 4.4 spikes s^{-1} in young monkeys ($n = 12$ neurons), and 25.3 ± 3.7 spikes s^{-1} in old monkeys (11 neurons; t -test, $P > 0.8$; significant age \times cell-type interaction in

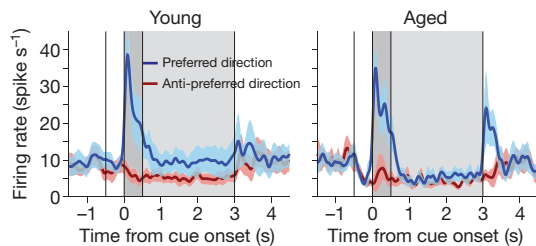


Figure 3 | Firing rates of dorsolateral PFC CUE cells remain stable in aged monkeys. The average firing rates of CUE cells in young monkeys (left graph; 10 neurons from a 7-year-old and 2 neurons from a 9-year-old monkey), did not differ from the firing rate in the oldest monkey (right graph; 11 neurons from a 21-year-old monkey), or from the firing rate averaged for both middle-aged (5 neurons from 13-year-old monkey, not shown separately) and old monkeys (t -test, $P > 0.7$). CUE cells may receive direct, 'bottom-up' excitation from parietal association cortex⁶, which may be less vulnerable to subtle molecular changes with advancing age. A subset of CUE cells recorded in the aged monkey showed high firing rates during the response period. Error envelope represents s.e.m.

two-way ANOVA, $P < 0.05$). These data indicate that reductions in memory-related firing rate do not arise from generalized changes with advancing age affecting all neurons but, rather, are especially evident in recurrent circuits that must maintain firing in the absence of 'bottom-up' sensory stimulation.

It is important to identify which changes in the ageing brain contribute to reduced firing during the delay period. There are many brain alterations associated with advancing age²⁷, including decreased PFC grey matter volume²⁸, focal changes in white matter²⁹ and dendritic spine loss³⁰, all of which correlate with cognitive decline. Importantly, spine loss is especially prominent in layer III—the layer where the recurrent excitatory networks reside—and thin-type spines are the most vulnerable in the aged PFC³⁰. Immunoelectron microscopy indicates that thin spines have the greatest concentration of cAMP–HCN-channel signalling proteins, indicating that disinhibition of cAMP

signalling with advancing age may weaken thin spines in particular⁹. Thus, we tested whether inhibition of cAMP signalling in the PFC could partially restore the working-memory-related firing of aged neurons, or whether reductions in firing were irreversible owing to immutable architectural changes in the aged brain. Drugs were applied near the recorded neurons using iontophoresis, whereby a small electrical current is applied to extrude charged molecules from glass pipettes attached to the recording electrode. Only a minute amount of drug was released, sufficient to alter the firing of nearby neurons, without altering behavioural performance.

Agents that inhibit cAMP signalling or block HCN or KCNQ channels restored persistent firing during the delay period of the working memory task (Fig. 4). For example, iontophoresis of the α_{2A} agonist, guanfacine (Fig. 4a, e), or the cAMP–PKA inhibitor, Rp-cAMPS (Fig. 4b, f), significantly increased firing during the delay period on trials when the cue had appeared at the neuron's preferred direction. In contrast, the PDE4 inhibitor, etazolate—which increases cAMP signalling—further decreased neuronal firing in aged neurons ($P < 0.001$; Supplementary Fig. 5). We also tested whether blockade of HCN or KCNQ channels could restore firing, given that cAMP–PKA signalling increases the open state of these ion channels. As shown in Fig. 4c and g, a low dose of the HCN channel blocker ZD7288 significantly enhanced the delay-related firing rates of neurons in aged monkeys. KCNQ channels were also of interest, as *in vitro* physiological characterizations of PFC neurons in aged primates have found increases in the slow afterhyperpolarization, which is mediated in part by KCNQ channels²⁶. As shown in Fig. 4d and h, blockade of KCNQ channels with XE991 increased delay-related firing in aged PFC neurons. Thus, agents that reduced cAMP opening of HCN or KCNQ channels all restored firing rates to levels resembling those observed in younger monkeys. These findings are consistent with behavioural data showing that guanfacine and Rp-cAMPS can enhance working memory performance in aged animals when administered systemically (guanfacine) or directly into the rat PFC (guanfacine or Rp-cAMPS)²² (see Supplementary Information). On the basis of these data, guanfacine

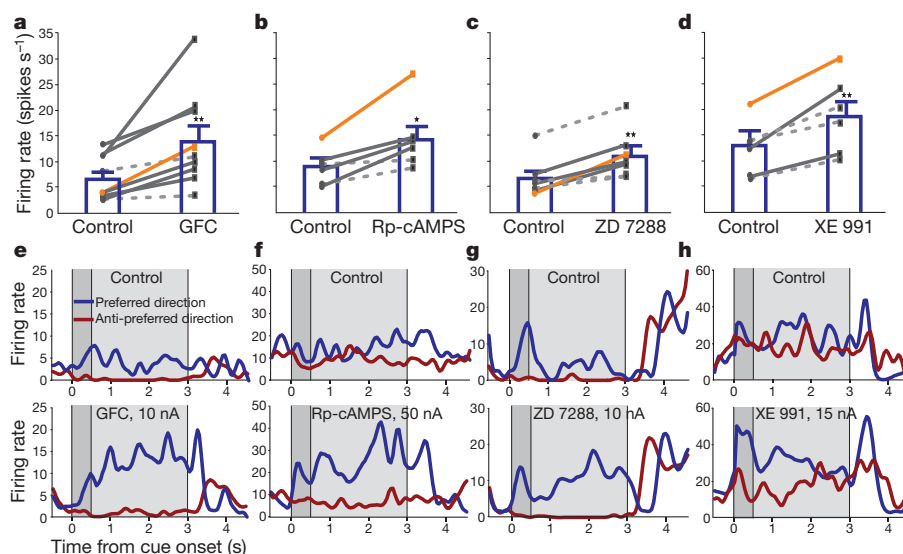


Figure 4 | Iontophoresis of compounds that inhibit cAMP–PKA signalling, or block HCN or KCNQ channel signalling, strengthens delay-related firing in aged PFC DELAY neurons. a–d, A summary of the results showing a significant increase in population-average firing rate for the neuron's preferred direction compared to control conditions (paired t -test, $P < 0.01$ for a, c and d, and $P < 0.05$ for b) following iontophoresis of the α_{2A} adrenergic agonist guanfacine applied at 10 nA (GFC; a, significant effects in 7 out of 9 neurons, t -test, $P < 0.05$, indicated by solid lines); the cAMP inhibitor Rp-cAMPS at 50 nA (b, significant in 4 out of 6 neurons); the HCN channel blocker, ZD7288 at 15 nA (c, significant in 4 out of 7 neurons); and the KCNQ channel blocker

XE991 at 15 nA (d, significant in 3 out of 6 neurons). In all cases, significant effects were found more frequently than expected by chance (binomial test, $P < 0.005$). The orange lines represent the individual neurons shown in e–h. e–h, Individual examples of neurons under control conditions (top) firing to their preferred (blue trace) or anti-preferred (red trace) directions, compared to their firing patterns following iontophoresis of guanfacine (e), Rp-cAMPS (f), ZD7288 (g) or XE991 (h). The orange lines in a–d indicate the individual neurons shown in e–h. Error bars are s.e.m. * $P < 0.05$, ** $P < 0.01$ significant difference between drug versus control for the neuron's preferred direction.

is currently being tested in elderly humans with PFC cognitive deficits (<http://www.clinicaltrials.gov>: trial number NCT00935493).

The current study revealed a physiological basis for age-related working memory decline in the primate brain, with a reduction of memory-related firing beginning in middle age and worsening with advancing age. This marked change in network physiology may render higher cortical circuits especially vulnerable to neurodegenerative processes such as Alzheimer's disease. However, these studies also uncovered more hopeful data showing that restitution of the proper neurochemical environment can partially restore physiological integrity. These data establish that cognitive changes with advancing age are malleable, and that there is potential to restore at least some cognitive abilities in the elderly. Maintaining strong PFC physiology into advanced age will be an important advantage in an increasingly complex, ageing society.

METHODS SUMMARY

All experiments were performed in accordance with National Institutes for Health guidelines for animal research and were approved by the Yale Institutional Animal Care and Use Committee.

Behavioural training. Monkeys were trained on the oculomotor delayed response (ODR) task (Fig. 1b) as reported elsewhere⁸, with special care to minimize all stress. The aged monkeys took longer to learn the task, rested more frequently than young monkeys during the testing sessions, and performed fewer trials each day than younger animals. A brief delay period (2.5 s) was chosen to ensure that all monkeys in the study performed above 85% correct during training, and could maintain high levels of performance during the study (Supplementary Fig. 6).

Physiology and iontophoresis. Single unit recordings were made from the dorsolateral PFC surrounding the caudal portion of the principal sulcus (Fig. 1a). The recording procedures and iontophoresis methods have been described previously⁸. Single neuron activity was readily isolated (Supplementary Fig. 7). Denoting the average activity during the 0.5 s cue period and that during the 2.5 s delay period during the trials in preferred directions as *C* and *D*, neurons were classified as CUE cells, when the ratio *D/C* was >0.5, and as DELAY cells otherwise. Statistical analyses, including ANOVA and regression analyses, were performed using Matlab (MathWorks). The effect of age on firing rate during a series of 0.5 s epochs was tested using a regression model, and its statistical significance with a *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2010; accepted 23 May 2011.

Published online 27 July 2011.

- West, R. L. An application of prefrontal cortex function theory to cognitive aging. *Psychol. Bull.* **120**, 272–292 (1996).
- Cabeza, R., Anderson, N. D., Houle, S., Mangels, J. A. & Nyberg, L. Age-related differences in neural activity during item and temporal-order memory retrieval: a positron emission tomography study. *J. Cogn. Neurosci.* **12**, 197–206 (2000).
- Gazzaley, A., Cooney, J. W., Rissman, J. & D'Esposito, M. Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neurosci.* **8**, 1298–1300 (2005).
- Prakash, R. S. *et al.* Age-related differences in the involvement of the prefrontal cortex in attentional control. *Brain Cogn.* **71**, 328–335 (2009).
- Goldman-Rakic, P. S. in *Handbook of Physiology, The Nervous System, Higher Functions of the Brain* Vol. 5 (ed. Plum, F.) 373–417 (American Physiological Society, 1987).
- Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
- Robbins, T. W. & Arnsten, A. F. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu. Rev. Neurosci.* **32**, 267–287 (2009).
- Wang, M. *et al.* α 2A-adrenoceptor stimulation strengthens working memory networks by inhibiting cAMP-HCN channel signaling in prefrontal cortex. *Cell* **129**, 397–410 (2007).

- Arnsten, A. F. T., Paspalas, C. D., Gamo, N. J., Yang, Y. & Wang, M. Dynamic network connectivity: a new form of neuroplasticity. *Trends Cogn. Sci.* **14**, 365–375 (2010).
- Bowles, R. P. & Salthouse, T. A. Assessing the age-related effects of proactive interference on working memory tasks using the Rasch model. *Psychol. Aging* **18**, 608–615 (2003).
- Royall, D. R., Palmer, R., Chiodo, L. K. & Polk, M. J. Normal rates of cognitive change in successful aging: the freedom house study. *J. Int. Neuropsychol. Soc.* **11**, 899–909 (2005).
- Burke, S. N. & Barnes, C. A. Neural plasticity in the ageing brain. *Nature Rev. Neurosci.* **7**, 30–40 (2006).
- Cappell, K. A., Gmeindl, L. & Reuter-Lorenz, P. A. Age differences in prefrontal recruitment during verbal working memory maintenance depend on memory load. *Cortex* **46**, 462–473 (2010).
- Davis, H. P. *et al.* Lexical priming deficits as a function of age. *Behav. Neurosci.* **104**, 288–297 (1990).
- Bucur, B. & Madden, D. J. Effects of adult age and blood pressure on executive function and speed of processing. *Exp. Aging Res.* **36**, 153–168 (2010).
- Sisodia, S. S., Martin, L. J., Walker, L. C., Borchelt, D. R. & Price, D. L. Cellular and molecular biology of Alzheimer's disease and animal models. *Neuroimaging Clin. N. Am.* **5**, 59–68 (1995).
- Moore, T. L., Killiany, R. J., Herndon, J. G., Rosene, D. L. & Moss, M. B. Executive system dysfunction occurs as early as middle-age in the rhesus monkey. *Neurobiol. Aging* **27**, 1484–1493 (2006).
- Rapp, P. R. & Amaral, D. G. Evidence for task-dependent memory dysfunction in the aged monkey. *J. Neurosci.* **9**, 3568–3576 (1989).
- Herndon, J. G., Moss, M. B., Rosene, D. L. & Killiany, R. J. Patterns of cognitive decline in aged rhesus monkeys. *Behav. Brain Res.* **87**, 25–34 (1997).
- Rypma, B. & D'Esposito, M. Isolating the neural mechanisms of age-related changes in human working memory. *Nature Neurosci.* **3**, 509–515 (2000).
- George, M. S., Abbott, L. F. & Siegelbaum, S. A. Hyperpolarization-activated HCN channels inhibit subthreshold EPSPs through voltage-dependent interactions with M-type K^+ channels. *Nature Neurosci.* **12**, 577–584 (2009).
- Ramos, B. *et al.* Dysregulation of protein kinase A signaling in the aged prefrontal cortex: new strategy for treating age-related cognitive decline. *Neuron* **40**, 835–845 (2003).
- Moore, T. L. *et al.* Cognitive impairment in aged rhesus monkeys associated with monoamine receptors in the prefrontal cortex. *Behav. Brain Res.* **160**, 208–221 (2005).
- Downs, J. L. *et al.* Orexin neuronal changes in the locus coeruleus of the aging rhesus macaque. *Neurobiol. Aging* **28**, 1286–1295 (2007).
- Schoenbaum, G., Setlow, B., Saddoris, M. P. & Gallagher, M. Encoding changes in orbitofrontal cortex in reversal-impaired aged rats. *J. Neurophysiol.* **95**, 1509–1517 (2006).
- Luebke, J. I. & Chang, Y. M. Effects of aging on the electrophysiological properties of layer 5 pyramidal cells in the monkey prefrontal cortex. *Neuroscience* **150**, 556–562 (2007).
- Luebke, J., Barbas, H. & Peters, A. Effects of normal aging on prefrontal area 46 in the rhesus monkey. *Brain Res. Rev.* **62**, 212–232 (2010).
- Alexander, G. E. *et al.* Age-related regional network of magnetic resonance imaging gray matter in the rhesus macaque. *J. Neurosci.* **28**, 2710–2718 (2008).
- Peters, A. *et al.* Neurobiological bases of age-related cognitive decline in the rhesus monkey. *J. Neuropathol. Exp. Neurol.* **55**, 861–874 (1996).
- Dumitriu, D. *et al.* Selective changes in thin spine density and morphology in monkey prefrontal cortex correlate with aging-related cognitive impairment. *J. Neurosci.* **30**, 7507–7515 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by PHS grant P01AG030004 from the National Institute on Aging. The authors would like to thank J. Thomas, L. Ciavarella, S. Johnson, B. Brunson and M. Horn for their assistance in making this work possible.

Author Contributions M.W. and X.-J.W., J.A.M., D.L. and A.F.T.A. designed the experiments. M.W. carried out all the physiology experiments, with the help of Y.Y., N.J.G., L.E.J. and J.A.M. Data analyses were performed by M.W., D.L., J.A.M. and M.L. Computational modelling was performed by X.-J.W. All authors participated in the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interest: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.F.T.A. (amy.arnsten@yale.edu).

METHODS

Oculomotor delayed response task. Studies were performed on four adult male rhesus monkeys (*Macaca mulatta*) trained on the spatial oculomotor delayed response (ODR) task as previously described⁸. This task requires the monkey to make a memory-guided saccade to a remembered visuospatial target. Each trial began when the subject fixated at the central spot for 0.5 s (fixation period). Subsequently, a cue was illuminated for 0.5 s at one of eight peripheral targets (cue period). After the cue was extinguished, a 2.5 s delay period followed. During the cue and delay periods, the monkey was required to maintain central fixation. At the end of the delay, the fixation spot was extinguished, instructing the monkey to make a memory-guided saccade to the previously cued location (saccade period). The monkey was rewarded with fruit juice immediately after every successful response. The position of the stimulus was randomized over trials such that it had to be remembered on a trial-by-trial basis. The subject's eye position was monitored with the ISCAN Eye Movement Monitoring System, and the ODR task was generated by the TEMPO real-time system (Reflective Computing). The aged monkeys took longer to learn the task, rested more frequently than young monkeys during the testing sessions, and performed fewer trials each day than younger animals. A brief delay period (2.5 s) was chosen to ensure that all monkeys in the study performed above 85% correct during training, and could maintain high levels of performance during the study (Supplementary Fig. 6).

Recording locus. Before recording, the animal underwent a magnetic resonance image (MRI) scan to obtain exact anatomical coordinates of brain structures, which guided the placement of the chronic recording chambers. MRI-compatible materials were used for the implant so that the position of the recording chambers could be confirmed by MRI after implantation. The site of recordings in the present study were located in an area ranging from 0–5 mm anterior to the caudal end of the principal sulcus and –2–2 mm medial to the principal sulcus.

Pharmacology, physiology and data acquisition. Guanfacine, XE991 and ZD7288 (Tocris) and etazolate (Sigma) were dissolved at 0.01 M in triple-distilled water (adjusted with HCl to pH 3.5–4.0). Rp-cAMP (Sigma) was dissolved at 0.01 M in triple-distilled water (adjusted with NaOH to pH 9).

Iontophoretic electrodes were constructed with a 20- μ m pitch carbon fibre (ELSI) inserted in the central barrel of a seven-barrel non-filamented capillary glass (Friedrich and Dimmock). The assembly was pulled using a multipipette electrode puller (MicroData Instrument) and the tip was bevelled to obtain the finished electrode. Finished electrodes had impedances of 0.3–1.0 M Ω (at 1 kHz) and tip sizes of 30–40 μ m. The outer barrels of the electrode were then filled with 3 drug solutions (two consecutive barrels each) and the solutions were pushed to the

tip of the electrode using compressed air. A Neurophore BH2 iontophoretic system (Medical Systems) was used to control of the delivery of the drugs. The drug was ejected at currents that varied from 5–50 nA. Retaining currents of –3 to –5 nA were used in a cyclical manner (1 s on, 1 s off) when not applying drugs. Drug ejection did not create noise in the recording, and there was no systematic change in either spike amplitude or time course at any ejection current.

The electrode was mounted on a MO-95 micromanipulator (Narishige) in a 25-gauge stainless steel guide tube. The dura was punctured using the guide tube to facilitate access of the electrode to cortex. Extracellular voltage was amplified using an AC/DC differential amplifier (A-M systems; Model 3000) and band-pass filtered (180 Hz–6 KHz, 20 dB gain, 4-pole Butterworth; Kron-Hite). Signals were digitized (20.83 kHz, micro 1401, Cambridge Electronics Design) and acquired using the Spike2 software (CED). Neural activity was analysed using waveform sorting by a template-matching algorithm, which made it possible to isolate more than one unit at the same recording site. Post-stimulus time histograms and rastergrams were constructed online to determine the relationship of unit activity to the task. Unit activity was measured in spikes per second. If the rastergrams displayed task-related activity, the units were recorded further and pharmacological testing was performed.

Data were first collected from the cell under a control condition in which at least eight trials at each of eight cue locations were obtained. On establishing the stability of the cells' activity, this control condition was followed by the drug application. Dose-dependent effects of the drug were tested in two or more consecutive conditions. Drugs were continuously applied at a relevant current throughout a given condition. Each condition had ~8 (typically 10 or more) trials at each location for statistical analyses of effects.

Data analysis. For purposes of data analysis, each trial in the ODR task was divided into four epochs: fixation, cue, delay and response (saccade). The fixation epoch lasted for 0.5 s. The cue epoch lasted for 0.5 s and corresponds to the stimulus presentation phase of the task. Delay lasted for 2.5 s and reflects the mnemonic component of the task. The response phase started immediately after the delay epoch and lasted ~1.5 s.

Two-way analysis of variance, ANOVA, was used to examine the spatial tuned task-related activity with regard to (1) different periods of the task (cue, delay, response versus fixation); and (2) different cue locations. One-way ANOVAs were used to assess the effect of the drug application on cells showing delay-related activity. Statistical analyses, including ANOVA and regression analyses were performed using Matlab (MathWorks). The effect of age on firing rate during a series of 0.5 s epochs was tested using a regression model, and its statistical significance with a *t*-test.

Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis

The International Multiple Sclerosis Genetics Consortium* & the Wellcome Trust Case Control Consortium 2*

Multiple sclerosis is a common disease of the central nervous system in which the interplay between inflammatory and neuro-degenerative processes typically results in intermittent neurological disturbance followed by progressive accumulation of disability¹. Epidemiological studies have shown that genetic factors are primarily responsible for the substantially increased frequency of the disease seen in the relatives of affected individuals^{2,3}, and systematic attempts to identify linkage in multiplex families have confirmed that variation within the major histocompatibility complex (MHC) exerts the greatest individual effect on risk⁴. Modestly powered genome-wide association studies (GWAS)^{5–10} have enabled more than 20 additional risk loci to be identified and have shown that multiple variants exerting modest individual effects have a key role in disease susceptibility¹¹. Most of the genetic architecture underlying susceptibility to the disease remains to be defined and is anticipated to require the analysis of sample sizes that are beyond the numbers currently available to individual research groups. In a collaborative GWAS involving 9,772 cases of European descent collected by 23 research groups working in 15 different countries, we have replicated almost all of the previously suggested associations and identified at least a further 29 novel susceptibility loci. Within the MHC we have refined the identity of the *HLA-DRB1* risk alleles and confirmed that variation in the *HLA-A* gene underlies the independent protective effect attributable to the class I region. Immunologically relevant genes are significantly overrepresented among those mapping close to the identified loci and particularly implicate T-helper-cell differentiation in the pathogenesis of multiple sclerosis.

We performed a large GWAS as part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) project. Cases were recruited through the International Multiple Sclerosis Genetics Consortium (IMSGC) and compared with the WTCCC2 common control set^{12,13} supplemented by data from the control arms of existing GWAS. We introduced a number of novel quality control methods for processing these data sets (see Supplementary Information), which ultimately provided reliable information from 9,772 cases and 17,376 controls (Fig. 1a). After single nucleotide polymorphism (SNP)-based quality controls, data from 465,434 autosomal SNPs, common to all internally and externally generated data sets, were available for analysis.

The multi-population nature of our study (Fig. 1a, b) afforded an opportunity to assess various published approaches for controlling the potential confounding effects of population structure, several of which (in the event) proved unhelpful (see Supplementary Information). Although not common in primary GWAS undertaken to date, the challenge of combining data across populations, in contexts where not all case samples have controls available from the same population (thus precluding standard meta-analytical techniques), may become more routine as study sizes increase.

We attempted analyses of the non-United Kingdom (UK) data with the now widespread technique of using principal components as covariates to correct for structure. However, even use of all seven top principal components that captured genome-wide effects in our data

resulted in an unacceptably high genomic inflation: for example, the genomic control factor¹⁴ (λ) was $\lambda = 1.2$. We tried to reduce the genomic inflation by discarding the case samples that seemed least well matched to control sets. Removal of half the available cases in this fashion only reduced λ to 1.1. In another approach to handling structure, statistical clustering algorithms were successful in identifying subgroups of the data within which cases and controls seemed well matched for ancestry (see Supplementary Fig. 17). However, tests within these subgroups combined via fixed-effects meta-analysis also yielded unacceptably high genomic inflation ($\lambda > 1.4$) in an analysis with seven matched subgroups of cases and controls. Lastly, we applied a novel variance components method (similar to one described previously¹⁵), separately to the UK and non-UK data sets, which explicitly accounts for correlations among the

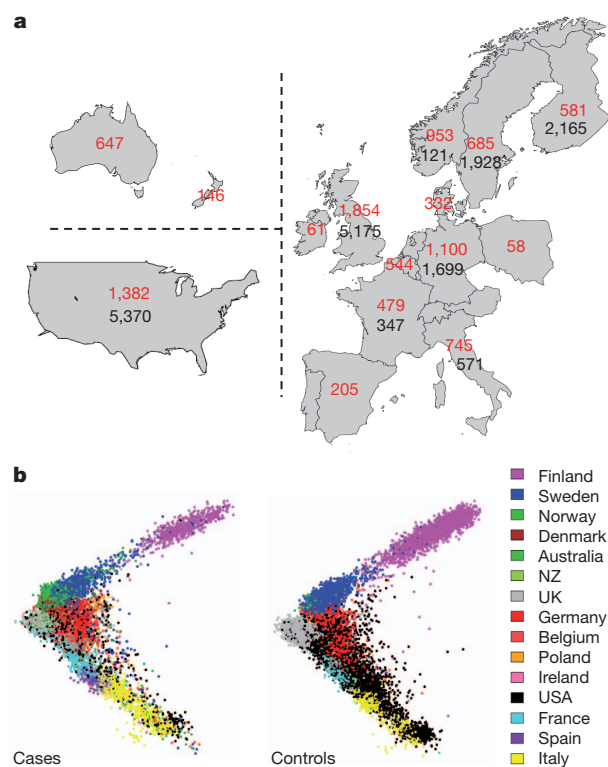


Figure 1 | Distribution of cases and controls. **a**, **b**, All cases and controls were drawn from populations with European ancestry; cases from 15 countries and controls from 8. **a**, Numbers of case (red) and control (black) samples from each country. **b**, The projection of samples onto the first two principal components of genetic variation, with cases shown on the left and controls on the right. The axes are orientated to approximate the geography, and samples are colour coded as indicated in the legend. NZ, New Zealand. We genotyped the cases (9,772) and some Swedish controls (527) using the Illumina Human 660-Quad platform, and the UK controls (5,175, the WTCCC2 common control set^{12,13}) using the Illumina 1.2M platform. All other controls were genotyped externally using various Illumina genotyping systems (see Supplementary Information).

*A list of authors and their affiliations appears at the end of the paper; membership of both consortia is listed in Supplementary Information.

phenotypes of individuals resulting from relatedness, allowing us to deal successfully with all sources of structure in our samples (see Supplementary Information for details of the linear mixed model we used). For example, the genomic inflation was reduced to $\lambda = 0.995$ in the UK and 1.016 in the non-UK data (see also Supplementary Information). After fixed-effects meta-analysis of the results from the UK and non-UK data sets, the inflation factor was $\lambda = 1.045$. We adopted this approach for all subsequent non-MHC association analyses.

Outside the MHC we identified 95 distinct regions having at least one SNP associated with multiple sclerosis at $P_{\text{GWAS}} < 1 \times 10^{-4.5}$; in six of these 95 regions conditional analysis revealed an additional SNP showing association to the same locus (one locus containing two such SNPs). In total we took all 102 SNPs forward to replication, which we performed using data from previously reported multiple sclerosis GWAS^{8,9} and the iControl database (excluding any WTCCC controls previously used in these studies). In total, the replication analysis included data from 4,218 cases and 7,296 controls. These were considered in six independent strata after which results were combined through a fixed-effects meta-analysis. For 98 of the 102 SNPs, the same

allele was overrepresented in cases compared to controls. Twenty three of the 26 previously known or strongly suggested multiple-sclerosis-associated loci were replicated in our primary GWAS with $P_{\text{GWAS}} < 1 \times 10^{-3}$. Our GWAS and replication also revealed another 29 novel associated regions (defined as having $P_{\text{GWAS}} < 1 \times 10^{-4.5}$, one-sided $P_{\text{replication}} < 0.05$ and $P_{\text{combined}} < 5 \times 10^{-8}$), and a further 5 regions with strong evidence for association (with $P_{\text{GWAS}} < 1 \times 10^{-4.5}$, one-sided $P_{\text{replication}} < 0.05$ and $P_{\text{combined}} < 5 \times 10^{-7}$). In one previously reported locus and two novel loci, additional SNPs were identified as being conditionally important in explaining risk. Just over one third of the identified loci overlap with regions already confirmed as associated with at least one other autoimmune disease (according to the GWAS catalogue, <http://www.genome.gov/gwastudies/>). Results for both the previously established and novel loci are shown in Fig. 2 and Supplementary Tables 1–3; and details of all 102 SNPs taken to replication are available in Supplementary Data.

To assess objectively the collective evidence across the associated regions for particular classes of genes, we performed statistical analyses to look for enrichment of genes with similar function. We first identified

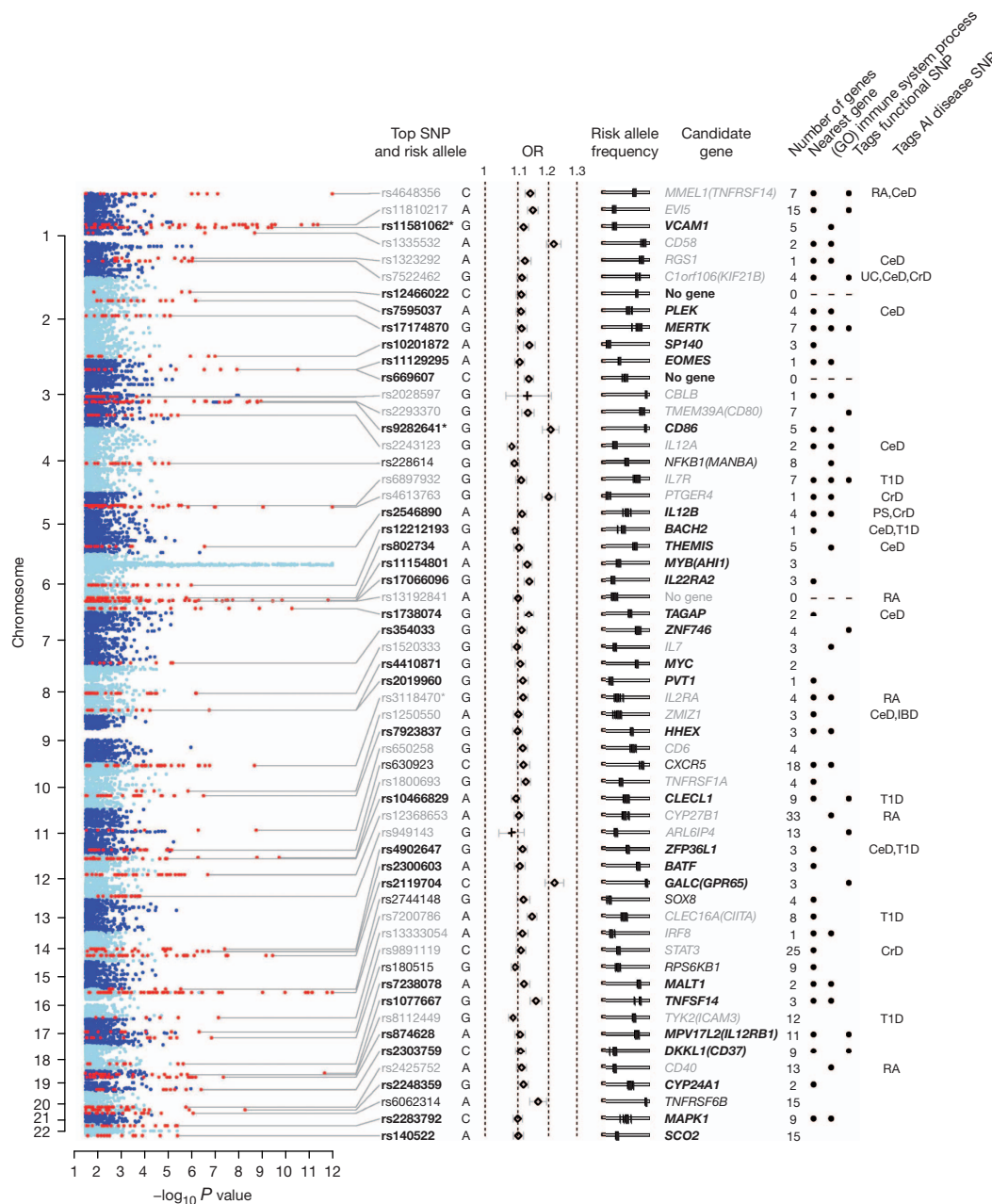


Figure 2 | Regions of the genome showing association to multiple sclerosis. Columns from left to right: first, evidence for association from the linear mixed model analysis of the discovery data (thresholded at $-\log_{10}P$ value = 12). Non-MHC regions containing associated SNPs are shown in red and are labelled with the rs number (bold for newly identified loci, black for strong evidence, grey for previously reported) and risk allele of the most significant SNP. Asterisk indicates that the locus contains a secondary SNP signal. Second, OR and 95% confidence intervals estimated from the meta-analysis of the discovery and replication data (+ indicates estimates for previously known loci from discovery data only). Third, risk allele frequency estimates in each of the control populations used in the study (each is shown as a vertical bar on a scale from 0 to 1 going left to right). For each region of association the number of genes is reported (fifth), and where non-zero a candidate gene is given (fourth). Black dots indicate that the candidate gene is physically the nearest gene (sixth) included in the 'immune system process' GO term (seventh). Eight, when the most-significant SNP tags an SNP predicted to have an impact on the function of the candidate gene this is indicated. Where such a SNP exists, the gene involved is selected as the candidate gene; otherwise the nearest gene is selected unless there are strong biological reasons for a different choice. The final column indicates SNPs that are correlated ($r^2 > 0.1$) with SNPs reported to be associated with other autoimmune (AI) diseases. CeD, coeliac disease; CrD, Crohn's disease; PS, psoriasis; RA, rheumatoid arthritis; T1D, type 1 diabetes; UC, ulcerative colitis. An interactive version of the figure is available at <http://www.well.ox.ac.uk/wtccc2/ms>.

the nearest gene to the lead SNP in each of the (52) regions of association and used the Gene Ontology (GO) database¹⁶ to define sets of functionally related genes (GO terms). We then tested whether the set of nearest genes was enriched for particular GO terms using Fisher's exact test. The GO terms having the most significant enrichment include genes linked to lymphocyte function ($P = 3.2 \times 10^{-11}$, odds ratio (OR) = 35.96) and in particular those with a role in T-cell activation and proliferation ($P = 1.85 \times 10^{-9}$, OR = 40.85). These are representative of a larger group associated with various components of the GO 'immune system process' ($P = 8.6 \times 10^{-11}$, OR = 9.12). A similar analysis based on all genes in or near association regions showed similar enrichment, as did independent analyses based on nearest gene or all genes in our next tier of signals, the 42 regions taken to replication but not meeting the thresholds above for association (see Supplementary Data). Although GO immune system genes only account for 7% of human genes, in 30% of our association regions the nearest gene to the lead SNP is an immune system gene. As an illustration, Fig. 3 shows a schematic of genes involved in the T-helper-cell differentiation pathway; a notable number show strong evidence for association with multiple sclerosis, particularly those acting as cell surface receptors. We infer from this pathway analysis of our GWAS signals that specific classes of immune system genes are especially important in the pathogenesis of multiple sclerosis.

Our screen not only implicates a multitude of genes coding for cytokine pathway (CXCR5, IL2RA, IL7R, IL7, IL12RB1, IL22RA2, IL12A, IL12B, IRF8, TNFRSF1A, TNFRSF14, TNFSF14), co-stimulatory (CD37, CD40, CD58, CD80, CD86, CLECL1) and signal transduction (CBLB, GPR65, MALTI, RGS1, STAT3, TAGAP, TYK2) molecules of immunological relevance, but also relates to previously reported environmental risk factors such as vitamin D^{9,17} (CYP27B1, CYP24A1) and therapies for multiple sclerosis including natalizumab¹⁸ (VCAM1) and daclizumab¹⁹ (IL2RA). There is a relative absence of genes relevant to potential pathways for neurodegeneration independent of inflammation (GALC, KIF21B).

To refine our understanding of the MHC associations in multiple sclerosis we imputed classical human leukocyte antigen (HLA) types at six loci (A, B, C, DQA1, DQB1 and DRB1)²⁰ and analysed these alongside the SNPs (see Supplementary Information for validation; at alleles responsible for the major signals described later, estimated specificity was at least 0.99 and sensitivity was at least 0.98, except for DRB1*13:03, where it was 0.88). Primary discovery was focused on the UK cohort with candidate signals being validated through support from additional case-control cohorts. Because of the extensive linkage disequilibrium within the MHC, we identified associated alleles in a stepwise

manner, selecting the most strongly associated to include in a general model, in turn, if $P_{UK} < 10^{-4}$ and $P_{combined} < 10^{-9}$ (Supplementary Information). At each stage we explored possible interactions and departures from the simple model in which risk increases multiplicatively with each additional copy of the relevant allele (additive increase on the log-odds scale) within the logistic risk framework.

Using this approach we found that DRB1*15:01 has the strongest association with multiple sclerosis among all classical and SNP alleles, with a consistent effect between cohorts ($P < 1 \times 10^{-320}$; Fig. 4a). The data are consistent with an additive effect on the log-odds scale for each additional allele. Conditioning on DRB1*15:01, we confirmed the presence of a protective class I allele and identified the signal as being driven by HLA-A*02:01 (as previously suggested²¹), with a consistent effect size across cohorts ($P = 9.1 \times 10^{-23}$; Fig. 4a). Again, we found no strong evidence for departure from additivity on the log-odds scale or statistical interaction with DRB1*15:01. Conditioning on both DRB1*15:01 and A*02:01 revealed additional risk associated with the strongly linked alleles DRB1*03:01 and DQB1*02:01 ($P = 3.6 \times 10^{-10}$; Fig. 4a; note that we cannot separate these alleles but for simplicity refer only to DRB1*03:01 later). Further conditioning identified an additional DRB1 risk allele DRB1*13:03 ($P = 1.3 \times 10^{-11}$; Fig. 4a). Although no other classical alleles meet the above criteria, we did observe several SNPs providing independent signals, the strongest coming from rs9277535_G (combined OR 1.28, $P = 2.2 \times 10^{-22}$), an allele known to be in linkage disequilibrium with DPB1*03:01 ($r^2 = 0.37$)²².

Analysis of the MHC SNP data using a genealogical method (GENECLUSTER)²³ offers an alternative means of relating our results to classical HLA alleles that provides additional insight into the underlying genetic architecture (see Supplementary Information). Figure 4b shows genealogical trees relating the classical alleles at DRB1 and HLA-A, together with the estimated evolutionary position of the mutations predicted by GENECLUSTER, as most completely modelling the association. At HLA-DRB1, three mutations are predicted, each of which implicates a clade of haplotypes carrying particular DRB1 alleles. All of the DRB1 alleles we have shown to be independently associated are included in these clades, each corresponding to a particular mutation. In addition, the analysis also explains why those haplotypes carrying the *08:01 allele have previously been shown to increase risk^{24,25} as they carry the same mutation as those bearing *13:03. At HLA-A, the predicted protective mutation is also concordant with our regression analysis of classical alleles in implicating *02:01 but, in addition, predicts that *68:01, *02:05 and *02:06 carry the same protective allele. All of these secondary predictions (increased risk from DRB1*08:01 and protection from HLA-A*68:01, *02:05 and *02:06) are supported in our regression

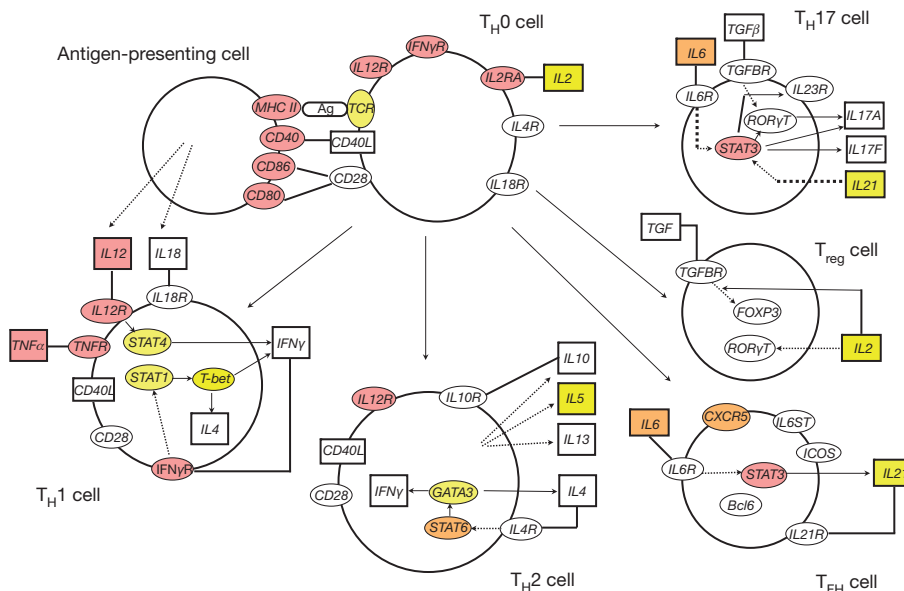


Figure 3 | Graphic representation of the T-helper-cell differentiation pathway. The figure is derived from an image generated by Ingenuity Pathway Analysis (IPA) software version 8.8 (Ingenuity Systems). Alphanumeric labels indicate the individual genes and gene complexes (nodes) included in the pathway (note that some are included more than once). Coloured nodes are those containing a gene implicated by proximity to an SNP showing evidence of association. Red, in bold or grey in Fig. 2 (plus MHC class II region and TNF); orange, other loci in Fig. 2 or discovery P value $< 1 \times 10^{-4.5}$ and consistent replication data; yellow, discovery P value $< 1 \times 10^{-3}$. Other molecules (proteins, vitamins etc) may also be of relevance in these processes but are not included here as they are not currently listed as being part of this particular pathway in the IPA database.

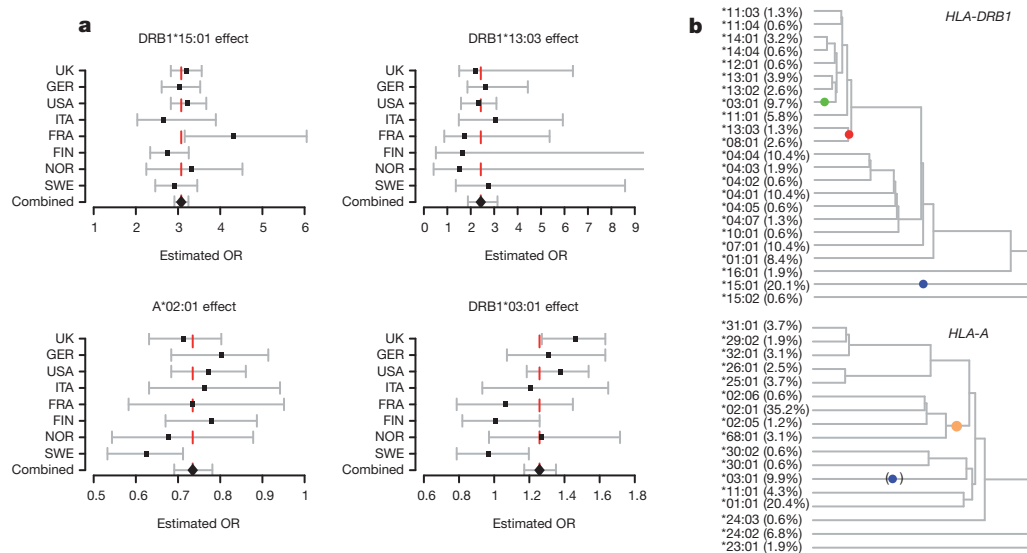


Figure 4 | Results for the main MHC alleles. **a**, Forest plots for each of the primary HLA alleles (HLA-A*02:01, DRB1*15:01, DRB1*03:01 and DRB1*13:03) showing consistency of effect across the populations and combined OR of 0.73, 3.1, 1.26 and 2.4, respectively (whiskers indicate 95% confidence intervals). **b**, The genealogical trees estimated for *DRB1* (top) and *HLA-A* (bottom). These trees were constructed using classical HLA and SNP typing data available from the HapMap CEU haplotype data. Each left-hand branch of the tree terminates on a set of haplotypes carrying a particular HLA allele. The coloured dots indicate the mostly likely locations for a disease-associated mutation as predicted by the GENECLUSTER program²³. In the

DRB1 tree, the blue dot captures a risk effect attributable to all haplotypes carrying the *15:01 allele. The green dot captures a risk effect carried by all haplotypes carrying the *03:01 allele and the red dot captures a risk effect on haplotypes carrying *13:03 or *08:01. In the *HLA-A* plot, the orange dot is a protective mutation lying at the root of all *02:01, *02:05, *02:06 and *68:01 alleles. The blue dot in brackets denotes a branch containing those *03:01 haplotypes that also carry DRB1*15:01; the GENECLUSTER prediction here is thus a reflection, due to linkage disequilibrium of the risk attributable to DRB1*15:01. The terminal branches are labelled with the allele carried by the haplotype and its frequency.

analysis of classical alleles but the power to detect them in the primary analyses is limited because each allele occurs at very low frequency.

We found no evidence for genetic associations with clinical course, severity of disease or month of birth, and no evidence of interaction with gender or DRB1*15:01 in any part of the genome (see Supplementary Information). However, analysis with respect to age at onset replicated the previously suggested association with the DRB1*15:01 allele²⁶. Although no other part of the genome contained individual SNPs showing strong evidence for association, risk alleles determining susceptibility are collectively more closely associated with age at onset than expected by chance, indicating that individual genetic susceptibility is inversely correlated with age at onset.

Our GWAS—large for any complex trait having a prevalence of 1:1,000 and involving diverse populations of European descent—has identified 29 novel susceptibility loci. Four mutations, one from class I and three from class II, with effects modelled in a simple multiplicative manner within and across loci are sufficient to account for most of the risk attributable to the MHC (see Supplementary Information). Although our data do not address the issue of which components within the nervous system are initially damaged by the inflammatory response, the overrepresentation of genes that influence T-cell maturation provides independent and compelling evidence that the critical disease mechanisms primarily involve immune dysregulation.

More generally, our study reinforces the view that the GWAS design, combined with very large experimental sample sizes and careful statistical analysis, provides valuable insights into the genetic architecture of common complex diseases. Here, this approach has identified many associated genetic variants close to genes, which are both individually interesting and collectively illuminate the roles of key biological pathways. It also provides indirect evidence that many more common variants of small effect contribute to genetic susceptibility for multiple sclerosis. Simple models, in which the previously known and newly identified variants affect risk multiplicatively, both within and across loci, explain a meaningful proportion (~20%, see Supplementary Information) of genetic risk for the disease. Important challenges lie ahead in understanding overlap between the genetic basis for

susceptibility in the context of different autoimmune diseases, and in uncovering the functional mechanisms underlying these associations.

METHODS SUMMARY

Details of case ascertainment, processing and genotyping, together with sample and genotyping quality control are provided in Supplementary Information. Statistical methods developed for testing the reliability of externally generated data sets, detecting samples with non-European ancestry, correcting for structure, classical HLA imputation and meta-analysis are also outlined in Supplementary Information. Results for all scans and all reported loci are described in detail in Supplementary Information.

Received 4 February; accepted 2 June 2011.

- Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502–1517 (2008).
- Dyment, D. A., Yee, I. M., Ebers, G. C. & Sadovnick, A. D. Multiple sclerosis in stepsiblings: recurrence risk and ascertainment. *J. Neurol. Neurosurg. Psychiatry* **77**, 258–259 (2006).
- Hemminki, K., Li, X., Sundquist, J., Hillert, J. & Sundquist, K. Risk for multiple sclerosis in relatives and spouses of patients diagnosed with autoimmune and related conditions. *Neurogenetics* **10**, 5–11 (2009).
- The International Multiple Sclerosis Genetics Consortium. A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* **77**, 454–467 (2005).
- The International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
- The Wellcome Trust Case Control Consortium & The Australo-Anglo-American Spondylitis Consortium. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nature Genet.* **39**, 1329–1337 (2007).
- Baranzini, S. E. *et al.* Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.* **18**, 767–778 (2009).
- De Jager, P. L. *et al.* Meta-analysis of genome scans and replication identify *CD6*, *IRF8* and *TNFRSF1A* as new multiple sclerosis susceptibility loci. *Nature Genet.* **41**, 776–782 (2009).
- The ANZgene Consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nature Genet.* **41**, 824–828 (2009).
- Sanna, S. *et al.* Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis. *Nature Genet.* **42**, 495–497 (2010).
- The International Multiple Sclerosis Genetics Consortium (IMSGC). Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am. J. Hum. Genet.* **86**, 621–625 (2010).
- The U.K. Parkinson's Disease Consortium & the Wellcome Trust Case Control Consortium 2. Dissection of the genetics of Parkinson's disease identifies an

- additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.* **20**, 345–353 (2011).
13. The Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nature Genet.* **42**, 985–990 (2010).
 14. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
 15. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348–354 (2010).
 16. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
 17. Pierrot-Deseilligny, C. & Souberbielle, J. C. Is hypovitaminosis D one of the environmental risk factors for multiple sclerosis? *Brain* **133**, 1869–1888 (2010).
 18. Steinman, L. A molecular trio in relapse and remission in multiple sclerosis. *Nature Rev. Immunol.* **9**, 440–447 (2009).
 19. Bielekova, B. *et al.* Effect of anti-CD25 antibody daclizumab in the inhibition of inflammation and stabilization of disease progression in multiple sclerosis. *Arch. Neurol.* **66**, 483–489 (2009).
 20. Leslie, S., Donnelly, P. & McVean, G. A statistical method for predicting classical HLA alleles from SNP data. *Am. J. Hum. Genet.* **82**, 48–56 (2008).
 21. Brynedal, B. *et al.* HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS ONE* **2**, e664 (2007).
 22. Field, J. *et al.* A polymorphism in the *HLA-DRB1* gene is associated with susceptibility to multiple sclerosis. *PLoS ONE* **5**, e13454 (2010).
 23. Su, Z., Cardin, N., Donnelly, P. & Marchini, J. A Bayesian method for detecting and characterizing allelic heterogeneity and boosting signals in genome-wide association studies. *Stat. Sci.* **24**, 430–450 (2009).
 24. Barcellos, L. F. *et al.* Heterogeneity at the *HLA-DRB1* locus and risk for multiple sclerosis. *Hum. Mol. Genet.* **15**, 2813–2824 (2006).
 25. Dymant, D. A. *et al.* Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum. Mol. Genet.* **14**, 2019–2026 (2005).
 26. Masterman, T. *et al.* HLA-DR15 is associated with lower age at onset in multiple sclerosis. *Ann. Neurol.* **48**, 211–219 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The principal funding for this study was provided by the Wellcome Trust (085475/B/08/Z, 085475/Z/08/Z, 075491/Z/04/Z and 068545/Z/02). The work was also supported by National Institutes of Health (AI076544, NS032830, NS049477, NS19142, NS049510, NS26799, NS43559, NS067305, CA104021, RR020092, RR024992 and K23N/S048869), US National Multiple Sclerosis Society (RG 4201-A-1), Nancy Davis Foundation, Cambridge NIHR Biomedical Research Centre, UK Medical Research Council (G0700061, G0000934), Multiple Sclerosis Society of Great Britain and Northern Ireland (898/08), Wolfson Royal Society Merit Award, Peter Doherty fellowship, Lagrange Fellowship, Harry Weaver Neuroscience Scholarships, Australian National Health and Medical Research Council (NHMC), Australian Research Council Linkage Program Grant, JHH Charitable Trust Fund, Multiple Sclerosis Research Australia, Health Research Council New Zealand, National MS Society of New Zealand, Wetenschappelijk Onderzoek Multiple Sclerose, Bayer Chair on Fundamental Genetic Research regarding the Neuroimmunological Aspects of Multiple Sclerosis, Biogen Idec Chair Translational Research in Multiple Sclerosis, FWO-Vlaanderen, Belgian Neurological Society, Danish Multiple Sclerosis Society, Neuropromise EU grant (LSHM-CT-2005-018637), Center of Excellence for Disease Genetics of the Academy of Finland, Sigrid Juselius Foundation, Helsinki University Central Hospital Research Foundation, Bundesministerium für Bildung und Technologie (KKNMS consortium Control MS), Deutsche Forschungsgemeinschaft, Institut National de la Santé et de la Recherche Médicale (INSERM), Association pour la Recherche sur la Sclérose En Plaques (ARSEP), Association Française contre les Myopathies (AFM), Italian Foundation for Multiple Sclerosis (FISM grants 2002/R/40, 2005/R/10, 2008/R/11 and 2008/R/15), Italian Ministry of Health (grant Giovani Ricercatori 2007 - D.lgs 502/92), Regione Piemonte (grants 2003, 2004, 2008, 2009), CRT Foundation, Turin, Moorfields/UCL Institute of Ophthalmology NIHR Biomedical Research Centre, Norwegian MS Register and Biobank, Research Council of Norway, South-Eastern and Western Norway regional Health Authorities, Ullevål University Hospital Scientific Advisory Council, Haukeland University Hospital, Amici Centro Sclerosi Multipla del San Raffaele (ACESM), Association of British Neurologists, Spanish Ministry of Health (FISPI060117), Bibbi and Niels Jensens Foundation, Montel Williams foundation, Hjärtfonden and Swedish medical research council (8691), Stockholm County Council (562183), Swedish Council for Working Life and Social Research, Gemeinnützige Hertie Stiftung, Northern California Kaiser Permanente members and Polpharma Foundation, and Washington University Institute of Clinical and Translational Sciences—Brain, Behavioral and Performance Unit. We acknowledge use of data from the British 1958 Birth Cohort, the UK National Blood Service, the popgen biobank, the KORA and MONICA Augsburg studies, the Accelerated Cure Project, the Brigham & Women's Hospital PhenoGenetic Project, the Swedish CAD project, the Norwegian Bone Marrow Donor Registry, the Children's Hospital of Philadelphia (CHOP), the Swedish Breast Cancer study, BRC-REFGENSEP (Pitié-Salpêtrière Centre d'Investigation Clinique (CIC) and Génomique) and HYPERGENES (HEALTH-F4-2007-201550). Projects received support from the German Ministry of Education and Research, the Helmholtz Zentrum München—National Research Center, the German National Genome Research Network (NGFN), the LMUinnovativ, the Knut and Alice Wallenberg Foundation, the Center for Applied Genomics from the Children's Hospital of Philadelphia Development Award, the Agency for Science & Technology and Research of Singapore, and the Susan G. Komen Breast Cancer Foundation. We thank S. Bertrand, J. Bryant, S. L. Clark, L. Collimedaglia,

G. Coniglio, J. S. Conquer, B. Colombo, T. Dibling, G. Eckstein, J. C. Eldred, G. Fischer, S. Gamble, P. Gregersen, R. Guerrero, C. Hind, P. Lichtner, L. Moiola, H. Mousavi, R. Naismith, R. J. Parks, R. Pearson, V. Pilato, M. Radaelli, E. Scarpini, C. R. Stribling, T. Strom, S. Taylor, D. Vukcevic and A. Wilk for their help and support. Detailed acknowledgements are available in Supplementary Information. This manuscript is dedicated to the memory of L. Peltonen, a member of both the IMSGC and the WTCCC2, in recognition of her contributions to, and her leadership in, human genetics.

Author Contributions Details of individual contributions are listed in Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (alastair.compston@medschl.cam.ac.uk; for the IMSGC) or P.D. (donnelly@well.ox.ac.uk; for the WTCCC2).

The International Multiple Sclerosis Consortium & the Wellcome Trust Case Control Consortium 2 (Membership of both consortia is listed in Supplementary Information.)

Stephen Sawcer^{1*}, Garrett Hellenthal^{2*}, Matti Pirinen^{2*}, Chris C. A. Spencer^{2*}, Nikolaos A. Patsopoulos^{3,4,5}, Loukas Moutsianas⁶, Alexander Dilthey⁶, Zhan Su², Colin Freeman², Sarah E. Hunt⁷, Sarah Edkins⁷, Emma Gray⁷, David R. Booth⁸, Simon C. Potter⁷, An Goris⁹, Gavin Band¹⁰, Annette Bang Oturai¹⁰, Amy Strange², Janna Saarela¹¹, Céline Bellenguez², Bertrand Fontaine¹², Matthew Gillman⁷, Bernhard Hemmer¹³, Rhian Williams⁷, Frauke Zipp^{14,15}, Alagurevathi Jayakumar⁷, Roland Martin¹⁶, Stephen Leslie¹⁷, Stanley Hawkins¹⁸, Eleni Giannoulakou², Sandra D'Alfonso¹⁹, Hannah Blackburn⁷, Filippo Martinelli Boneschi²⁰, Jennifer Liddle⁷, Hanne F. Harbo^{21,22}, Marc L. Perez⁷, Anne Spurkland²³, Matthew J. Waller⁷, Marcin P. Mycko²⁴, Michelle Ricketts⁷, Manuel Comabella²⁵, Naomi Hammond⁷, Ingrid Kockum²⁶, Owen T. McCann⁷, Maria Ban¹, Pamela Whittaker⁷, Anu Kempainen¹, Paul Weston⁷, Clive Hawkins²⁷, Sara Widaa⁷, John Zajicek²⁸, Serge Dronov⁷, Neil Robertson²⁹, Suzannah J. Bumpstead⁷, Lisa F. Barcellos^{30,31}, Rathi Ravindrarajah⁷, Roby Abraham²⁷, Lars Alfredsson³², Kristin Ardlie⁴, Cristin Aubin⁴, Amie Baker¹, Katharine Baker²⁹, Sergio E. Baranzini³³, Laura Bergamaschi¹⁹, Roberto Bergamaschi³⁴, Allan Bernstein³¹, Achim Berthele³⁵, Mike Boggild³⁵, Jonathan P. Bradfield³⁶, David Brassat³⁷, Simon A. Broadley³⁸, Dorothea Buck¹³, Helmut Butzkueven^{39,40,41,42}, Ruggero Capra⁴³, William M. Carroll⁴⁴, Paola Cavalla⁴⁵, Elisabeth G. Celius²¹, Sabine Cepok¹³, Rosetta Chiavacci³⁶, Françoise Clerget-Darpoux⁴⁶, Kathleen Clynsters⁹, Giancarlo Comi²⁰, Mark Cossumb²⁹, Isabelle Courou-Rebeix¹², Mathew B. Cox⁴⁷, Wendy Cozen⁴⁸, Bruce A. C. Cree³³, Anne H. Cross⁴⁹, Daniele Cusi⁵⁰, Mark J. Daly^{4,51,52}, Emma Davis⁵³, Paul I. W. de Bakker^{3,4,54,55}, Marc Debouverie⁵⁶, Marie Beatrice D'hooghe⁵⁷, Katherine Dixon⁵⁸, Rita Dobosi⁹, Bénédicte Dubois⁹, David Ellinghaus⁵⁸, Irina Elovaaara^{59,60}, Federica Esposito⁶², Claire Fontenille¹², Simon Foote⁶¹, Andre Franke⁵⁸, Daniela Galimberti⁶², Angelo Ghezzi⁶³, Joseph Glessner³⁶, Refujia Gomez³³, Olivier Gout⁶⁴, Colin Graham⁶⁵, Struan F. A. Grant^{36,66,67}, Franca Rosa Guerin⁶⁸, Hakon Hakonarson^{36,66,67}, Per Hall⁶⁹, Anders Hamsten⁷⁰, Hans-Peter Hartung⁷¹, Rob N. Heard⁸, Simon Heath⁷², Jeremy Hobart²⁸, Muna Hoshi¹³, Carmen Infante-Duarte⁷³, Gillian Ingram²⁹, Wendy Ingram²⁸, Talat Islam⁴⁸, Maja Jagodic²⁶, Michael Kabesch⁷⁴, Allan G. Kermod⁴⁴, Trevor J. Kilpatrick^{39,40,75}, Cecilia Kim³⁶, Norman Klopp⁷⁶, Keijo Koivisto⁷⁷, Malin Larsson⁷⁰, Mark Lathrop⁷², Jeannette S. Lechner-Scott^{77,78}, Maurizio A. Leone⁷⁹, Virpi Leppä^{11,80}, Ulrika Liljedahl⁸¹, Izaura Lima Bomfim⁸², Robin R. Lincoln³³, Jenny Link²⁶, Jianjun Liu⁸², Åslaug R. Lorentzen^{22,83}, Sara Lupoli^{50,84}, Fabio Macciardi^{50,85}, Thomas Mack⁴⁸, Mark Marriott^{39,40}, Vittorio Martinelli²⁰, Deborah Mason⁸⁶, Jacob L. McCauley⁸⁷, Frank Mentch³⁶, Inger-Lise Mero^{21,83}, Tania Mihalova²⁷, Xavier Montalban²⁵, John Mottershead^{88,89}, Kjell-Morten Myhr^{90,91}, Paola Naldi⁷⁹, William Ollier⁵³, Alison Page⁹², Aarno Palotie^{7,11,93,94}, Jean Pelletier⁹⁵, Laura Piccio⁴⁹, Trevor Pickersgill²⁹, Fredrik Piehl²⁶, Susan Pobylawoj⁹, Hong L. Quach³⁰, Patricia P. Ramsay³⁰, Mauri Reunanen⁹⁶, Richard Reynolds⁹⁷, John D. Rioux⁹⁸, Mariaemma Rodegher²⁰, Sabine Roesner¹⁶, Justin P. Rubio³⁹, Ina-Maria Rückert⁷⁶, Marco Salvetti⁹⁹, Erika Salvi^{50,100}, Adam Santaniello³³, Catherine A. Schaefer³¹, Stefan Schreiber^{58,101}, Christian Schulze¹⁰², Rodney J. Scott⁴⁷, Finn Sellebjerg¹⁰, Krzysztof W. Selma²⁴, David Sexton¹⁰³, Ling Shen³¹, Brigid Simms-Acuna³¹, Sheila Skidmore¹, Patrick M. A. Sleiman^{36,66}, Cathrine Smead²¹, Per Soelberg Sørensen¹⁰, Helle Bach Søndergaard¹⁰, Jim Stankovich⁶¹, Richard C. Strange²⁷, Anna-Maija Sulonen^{11,80}, Emilie Sundqvist²⁶, Ann-Christine Syvänen⁵¹, Francesca Taddeo¹⁰⁰, Bruce Taylor⁶¹, Jenefer M. Blackwell^{104,105}, Pentti Tienari¹⁰⁶, Elvira Bramon¹⁰⁷, Ayman Tourbah¹⁰⁸, Matthew A. Brown¹⁰⁹, Ewa Tronczynska²⁴, Juan P. Casas¹¹⁰, Niall Tubridy^{40,111}, Aiden Corvin¹¹², Jane Vickery²⁸, Janusz Jankowski¹¹³, Pablo Villoslada¹¹⁴, Hugh S. Markus¹¹⁵, Kai Wang^{36,66}, Christopher G. Mathew¹¹⁶, James Wason¹¹⁷, Colin N. A. Palmer¹¹⁸, H-Erich Wichmann^{76,119,120}, Robert Plomin¹²¹, Ernest Willoughby¹²², Anna Rautanen², Juliane Winkelmann^{13,123,124}, Michael Wittig^{58,125}, Richard C. Trembath¹¹⁶, Jacqueline Yaouanq¹²⁶, Ananth C. Viswanathan¹²⁷, Haitao Zhang^{36,66}, Nicholas W. Wood¹²⁸, Rebecca Zuvich¹⁰³, Panos Deloukas⁷, Cordelia Langford⁷, Audrey Duncanson¹²⁹, Jorge R. Oksenberg³³, Margaret A. Pericak-Vance⁸⁷, Jonathan L. Haines¹⁰³, Tomas Olsson²⁶, Jan Hillert²⁶, Adrian J. Iverson^{51,130}, Philip L. De Jager^{4,5,51}, Leena Peltonen⁷, Graeme J. Stewart⁸, David A. Hafler^{4,131}, Stephen L. Hauser³³, Gil McVean², Peter Donnelly^{2,6*} & Alastair Compston^{1*}

¹University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, BOX 165, Hills Road, Cambridge CB2 0QQ, UK. ²Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ³Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston,

- Massachusetts 02115, USA. ⁴Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, 02142 Massachusetts, USA. ⁵Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁶Department of Statistics, University of Oxford, Oxford OX1 3TG, UK. ⁷Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ⁸Westmead Millennium Institute, University of Sydney, New South Wales 2145, Australia. ⁹Laboratory for Neuroimmunology, Section of Experimental Neurology, Katholieke Universiteit Leuven, 3000 Leuven, Belgium. ¹⁰Danish Multiple Sclerosis Center, Department of Neurology, Copenhagen University Hospital, Rigshospitalet, 2100 Copenhagen, Denmark. ¹¹Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki 00290, Finland. ¹²INSERM UMR S 975 CRICM, UPMC, Département de neurologie Pitié-Salpêtrière, AP-HP, 75013 Paris, France. ¹³Department of Neurology, Klinikum Rechts der Isar, Technische Universität München, Ismaninger Strasse 22, 81675 Munich, Germany. ¹⁴Department of Neurology, University Medicine Mainz, Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany. ¹⁵Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany. ¹⁶Institute for Neuroimmunology and Clinical MS Research (inims), Centre for Molecular Neurobiology, Falkenried 94, D-20251 Hamburg, Germany. ¹⁷Department of Clinical Pharmacology, University of Oxford, Old Road Campus Research Building, Old Road Campus, Oxford OX3 7DQ, UK. ¹⁸Queen's University Belfast, University Road, Belfast BT7 1NN, Northern Ireland, UK. ¹⁹Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, 28100 Novara, Italy. ²⁰Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy. ²¹Department of Neurology, Oslo University Hospital, N-0407 Oslo, Norway. ²²Department of Neurology, University of Oslo, N-0318 Oslo, Norway. ²³Institute of Basal Medical Sciences, University of Oslo, N-0317 Oslo, Norway. ²⁴Department of Neurology, Laboratory of Neuroimmunology, Medical University of Lodz, Kopcinskiego 22, 90-153 Lodz, Poland. ²⁵Clinical Neuroimmunology Unit, Multiple Sclerosis Center of Catalonia (CEM-Cat), Vall d'Hebron University Hospital, Barcelona 08035, Spain. ²⁶Department of Clinical Neurosciences, Centre for Molecular Medicine CMM, L8:04, Karolinska Institutet, Karolinska University Hospital, SE-171 76 Stockholm, Sweden. ²⁷Keele University Medical School, Stoke-on-Trent ST4 7NY, UK. ²⁸Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth, Clinical Neurology Research Group, Tamar Science Park, Plymouth PL6 8BX, UK. ²⁹Department of Neurology, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK. ³⁰Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, California 94720-7356, USA. ³¹Kaiser Permanente Northern California Division of Research, 2000 Broadway, Oakland, California 94612, USA. ³²Institute of Environmental Medicine, Karolinska Institutet, Box 210, 171 77 Stockholm, Sweden. ³³Department of Neurology, University of California San Francisco, 505 Parnassus Avenue, S-256, San Francisco, California 94143-0435, USA. ³⁴Neurological Institute C. Mondino, IRCCS, 27100 Pavia, Italy. ³⁵The Walton Centre for Neurology and Neurosurgery, Liverpool L7 9LJ, UK. ³⁶Center for Applied Genomics, The Children's Hospital of Philadelphia, 3615 Civic Center Blvd., Philadelphia, Pennsylvania 19104, USA. ³⁷INSERM U 1043 et Pôle Neurosciences, Hôpital Purpan, 31059 Toulouse, France. ³⁸School of Medicine, Griffith University, 4222, Australia. ³⁹Florey Neuroscience Institutes, University of Melbourne, Victoria 3010, Australia. ⁴⁰Royal Melbourne Hospital, Parkville, Victoria 3050, Australia. ⁴¹Box Hill Hospital, Monash University, Box Hill 3128, Australia. ⁴²Department of Medicine, RMH Cluster, University of Melbourne, Victoria 3010, Australia. ⁴³Multiple Sclerosis Centre, Department of Neurology, Ospedale Civili di Brescia, 25018 Brescia, Italy. ⁴⁴Centre for Neuromuscular and Neurological Disorders, University of Western Australia, Perth, Western Australia 6009, Australia. ⁴⁵Department of Neurosciences, University of Turin, A.O.U. San Giovanni Battista, 10126 Turin, Italy. ⁴⁶INSERM U669, Univ Paris-Sud, 94800 Villejuif, France. ⁴⁷University of Newcastle, University Drive, Callaghan, New South Wales 2308, Australia. ⁴⁸Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, 1441 Eastlake Ave. Room 4453, Los Angeles, California 90033, USA. ⁴⁹Department of Neurology, Washington University, St Louis, Missouri 63110, USA. ⁵⁰University of Milan, Department of Medicine, Surgery and Dentistry, AO San Paolo, University of Milan, c/o Filarete Foundation, Viale Ortes 22/4, 20139 Milano, Italy. ⁵¹Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵²Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵³The UK DNA Banking Network, Centre for Integrated Genomic Medical Research, University of Manchester M13 9PT, UK. ⁵⁴Department of Medical Genetics, Division of Biomedical Genetics, University Medical Center Utrecht, 3508 GA, Utrecht, The Netherlands. ⁵⁵Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508 GA, Utrecht, The Netherlands. ⁵⁶Service de Neurologie, Hôpital Central, 54035 Nancy, France. ⁵⁷National Multiple Sclerosis Center, 1820 Melsbroek, Belgium. ⁵⁸Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, D-24105, Germany. ⁵⁹Department of Neurology, Tampere University Hospital, FIN-33014 Tampere, Finland. ⁶⁰University of Tampere, Medical School, Tampere 33014, Finland. ⁶¹Menzies Research Institute Tasmania, University of Tasmania, Private Bag 23, Hobart, Tasmania 7000, Australia. ⁶²Department of Neurological Sciences, Centro Dino Ferrari, University of Milan, Fondazione Cà Granda, Ospedale Maggiore Policlinico, 20122 Milan, Italy. ⁶³Centro Studi Sclerosi Multipla, Ospedale di Gallarate, 21013 Gallarate (VA), Italy. ⁶⁴Service de Neurologie, Fondation Ophtalmologique Adolphe de Rothschild, 75019 Paris, France. ⁶⁵Belfast Health and Social Care Trust, City Hospital, Belfast BT9 7AB, Northern Ireland, UK. ⁶⁶Division of Genetics, The Children's Hospital of Philadelphia, 3615 Civic Center Blvd., Philadelphia, Pennsylvania 19104, USA. ⁶⁷Department of Pediatrics, University of Pennsylvania School of Medicine, 3615 Civic Center Blvd., Philadelphia, Pennsylvania 19104, USA. ⁶⁸Laboratory of Molecular Medicine and Biotechnology, Don C. Gnocchi Foundation IRCCS, S. Maria Nascente, 20148 Milan, Italy. ⁶⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institute, 17177 Stockholm, Sweden. ⁷⁰Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Center for Molecular Medicine, L8:03, Karolinska University Hospital Solna, S-171 76 Stockholm, Sweden. ⁷¹Department of Neurology, Heinrich-Heine-University, D-40225, Düsseldorf, Germany. ⁷²Commissariat à l'énergie atomique, Institut de Génétique, Centre National de Genotypage, 2 rue Gaston Cremieux, CP 5721, 91057 Evry Cedex, France. ⁷³Experimental and Clinical Research Center, Charité, Universitätsmedizin Berlin and Max Delbrück Center for Molecular Medicine, Berlin 13125, Germany. ⁷⁴Department for Paediatric Pneumology, Allergy and Neonatology, Hannover Medical School, Carl-Neuberg Strasse 1, D30625 Hannover, Germany. ⁷⁵Centre for Neuroscience, University of Melbourne, Victoria 3010, Australia. ⁷⁶Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Munich, Germany. ⁷⁷Seinäjo Central Hospital, Seinäjoki, 60220, Finland. ⁷⁸Hunter Medical Research Institute, John Hunter Hospital, Lookout Road, New Lambton, New South Wales 2305, Australia. ⁷⁹SCDU Neurology, Maggiore della Carità Hospital, 28100 Novara, Italy. ⁸⁰Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki 00290, Finland. ⁸¹Molecular Medicine and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Entrance 70, 3rd Floor, Res Dept 2, Univeristy Hospital, S-75185 Uppsala, Sweden. ⁸²Human Genetics and Cancer Biology, Genome Institute of Singapore, Singapore 138672. ⁸³Institute of Immunology, Oslo University Hospital, N-0027 Oslo, Norway. ⁸⁴Institute of Experimental Neurology (INSPE), San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy. ⁸⁵Department of Psychiatry and Human Behavior, University of California, Irvine (UCI), 5251 California Av. S.te 240, Irvine, California 92617, USA. ⁸⁶Christchurch School of Medicine, University of Otago, Christchurch 8041, New Zealand. ⁸⁷John P. Hussman Institute for Human Genomics and The Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, 1501 NW 10th Avenue, Miami, Florida 33136, USA. ⁸⁸Greater Manchester Centre for Clinical Neurosciences, Hope Hospital, Salford M6 8HD, UK. ⁸⁹The Department of Neurology, Dunedin Public Hospital, Otago 9016, New Zealand. ⁹⁰The Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University Hospital, N-5021 Bergen, Norway. ⁹¹Department of Clinical Medicine, University of Bergen, N-5021 Bergen, Norway. ⁹²Plymouth Hospitals NHS Trust, Department of Neurology, Derriford Hospital, Plymouth PL6 8DH, UK. ⁹³Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki 00014, Finland. ⁹⁴Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁹⁵Pôle Neurosciences Cliniques, Service de Neurologie, Hôpital de la Timone, 13005 Marseille, France. ⁹⁶Department Neurology, Oulu University Hospital, Oulu 90029, Finland. ⁹⁷UK MS Tissue Bank, Wolfson Neuroscience Laboratories, Imperial College London, Hammersmith Hospital, London W12 0NN, UK. ⁹⁸Université de Montréal & Montreal Heart Institute, Research Center, 5000 rue Belanger, Montreal, Quebec H1T 1C8, Canada. ⁹⁹Neurology and Center for Experimental Neurological Therapy (CENTERS), Sapienza University of Rome, 00189-Rome, Italy. ¹⁰⁰KOS Genetic Srl, Via Podgora, 7, 20123 Milan, Italy. ¹⁰¹Department of General Internal Medicine, University Hospital, Schleswig-Holstein, Christian-Albrechts-University, Kiel 24105, Germany. ¹⁰²Systems Biology and Protein-Protein Interaction, Center for Molecular Neurobiology, Falkenried 94, D-20251 Hamburg, Germany. ¹⁰³Center for Human Genetics Research, Vanderbilt University Medical Center, 519 Light Hall, Nashville, Tennessee 37232, USA. ¹⁰⁴Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, 100 Roberts Road, Subiaco, Western Australia 6008, Australia. ¹⁰⁵Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Cambridge CB2 0XY, UK. ¹⁰⁶Department of Neurology, Helsinki University Central Hospital and Molecular Neurology Programme, Biomedicum, University of Helsinki, FIN-00290 Helsinki, Finland. ¹⁰⁷Division of Psychological Medicine and Psychiatry, Biomedical Research Centre for Mental Health at the Institute of Psychiatry, King's College London and The South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AF, UK. ¹⁰⁸Service de Neurologie et Faculté de Médecine de Reims, Université de Reims Champagne-Ardenne, 51100 Reims, France. ¹⁰⁹University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia. ¹¹⁰Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. ¹¹¹St Vincent's University Hospital, Dublin 4, Ireland. ¹¹²Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland. ¹¹³Centre for Gastroenterology, Bart's and the London School of Medicine and Dentistry, London E1 2AT, UK. ¹¹⁴Department of Neurosciences, Institute of Biomedical Research August Pi Sunyer (IDIBAPS), Hospital Clinic of Barcelona, 08036, Spain. ¹¹⁵Clinical Neurosciences, St George's University of London, London SW17 0RE, UK. ¹¹⁶Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London SE1 9RT, UK. ¹¹⁷Medical Research Council Biostatistics Unit, Robinson Way, Cambridge CB2 0SR, UK. ¹¹⁸Biomedical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK. ¹¹⁹Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany. ¹²⁰Klinikum Grosshadern, Munich 81377, Germany. ¹²¹King's College London, Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London SE5 8AF, UK. ¹²²Department of Neurology, Auckland City Hospital, Grafton Road, Auckland 1010, New Zealand. ¹²³Institut für Humangenetik, Technische Universität München, 81675 Munich, Germany. ¹²⁴Institut für Humangenetik, Helmholtz Zentrum München, 85764 Neuherberg, Munich, Germany. ¹²⁵Pöppgen Biobank, Christian-Albrechts University Kiel, Kiel 24105, Germany. ¹²⁶Pôle Recherche et Santé Publique, CHU Pontchaillou, 35033 Rennes, France. ¹²⁷NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London EC1V 2PD, UK. ¹²⁸Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N 3BG, UK. ¹²⁹Molecular and Physiological Sciences, The Wellcome Trust, London NW1 2BE, UK. ¹³⁰Harvard NeuroDiscovery Center, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹³¹Department of Neurology & Immunology, Yale University Medical School, New Haven, 06520 Connecticut, USA.

*These authors contributed equally to this work.

†Deceased.

Induction of human neuronal cells by defined transcription factors

Zhiping P. Pang^{1*}, Nan Yang^{2*}, Thomas Vierbuchen^{2,3*}, Austin Ostermeier^{2,3}, Daniel R. Fuentes², Troy Q. Yang², Ami Citri⁴, Vittorio Sebastiano², Samuele Marro², Thomas C. Südhof^{1,5} & Marius Wernig^{2,3}

Somatic cell nuclear transfer, cell fusion, or expression of lineage-specific factors have been shown to induce cell-fate changes in diverse somatic cell types^{1–12}. We recently observed that forced expression of a combination of three transcription factors, *Brn2* (also known as *Pou3f2*), *Ascl1* and *Myt1l*, can efficiently convert mouse fibroblasts into functional induced neuronal (iN) cells¹³. Here we show that the same three factors can generate functional neurons from human pluripotent stem cells as early as 6 days after transgene activation. When combined with the basic helix–loop–helix transcription factor *NeuroD1*, these factors could also convert fetal and postnatal human fibroblasts into iN cells showing typical neuronal morphologies and expressing multiple neuronal markers, even after downregulation of the exogenous transcription factors. Importantly, the vast majority of human iN cells were able to generate action potentials and many matured to receive synaptic contacts when co-cultured with primary mouse cortical neurons. Our data demonstrate that non-neural human somatic cells, as well as pluripotent stem cells, can be converted directly into neurons by lineage-determining transcription factors. These methods may facilitate robust generation of patient-specific human neurons for *in vitro* disease modelling or future applications in regenerative medicine.

Encouraged by our recent findings in mouse cells¹³, we explored in this study whether human fibroblasts could also be directly induced to become functional neurons. This was unclear given the differences in the gene regulatory networks governing human and rodent neural development^{14–16}. First, we sought to determine whether forced expression of transcription factors could induce a neuronal fate in human pluripotent cells. To that end, we infected undifferentiated human embryonic stem (ES) cells in chemically defined N3 media¹⁷ with *Brn2*, *Ascl1* and *Myt1l* (BAM) using doxycycline (dox)-inducible lentiviral vectors together with an EGFP (enhanced green fluorescent protein) virus. The majority of ES cells were EGFP-positive 24 h after addition of doxycycline (Supplementary Fig. 1). Strikingly, as early as 3 days after doxycycline treatment, we observed bipolar neuron-like cells surrounding nearly all ES cell colonies (Fig. 1a and Supplementary Fig. 1). By day 8, cells with more mature neuronal morphologies that expressed both β -III-tubulin (Tuj1, also known as TUBB3) and MAP2 had migrated away from ES cell colonies and were present throughout the plate (Fig. 1b, c). In contrast, after infection with EGFP virus alone, no neuronal cells were generated during the same timeframe, and nearly all ES cells had died due to the cytosine β -D-arabino-furanoside (Ara-C) treatment. We then determined the relative contribution of the three factors and found that *Ascl1* alone was sufficient to induce MAP2-positive cells (Supplementary Fig. 2). The addition of *Brn2* or *Myt1l* or both did not increase the efficiency of neuronal differentiation but induced more complex morphologies. Cells infected with all three factors together had the most mature neuronal morphologies

(Supplementary Fig. 2). Electrophysiological analysis surprisingly revealed that as early as 6 days after induction all recorded cells ($n = 16$) generated action potentials (Fig. 1d). At day 15 after doxycycline treatment, the average resting membrane potential of neuronal cells was -51 ± 1.8 mV (mean \pm s.e.m., $n = 18$) (Fig. 1f and Supplementary Table 1). These ES-iN cells exhibited prominent

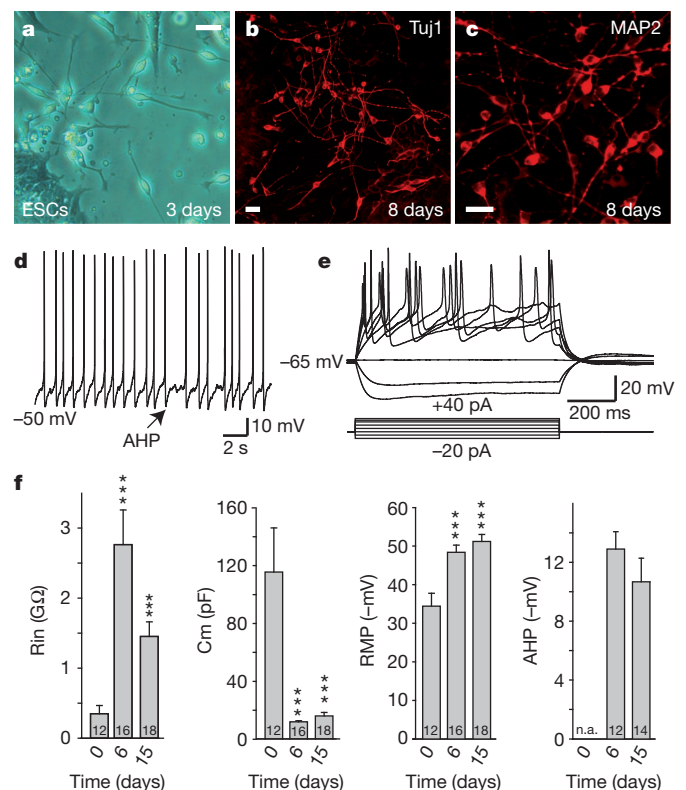


Figure 1 | Rapid generation of functional neurons from human ES cells.

a, Three days after induction, ES-iN cells showed bipolar neuronal morphologies. **b, c**, Eight days after induction, ES-iN cells expressed Tuj1 (**b**) and MAP2 (**c**). **d**, Spontaneous action potentials presumably caused by membrane potential fluctuations recorded from an ES-iN cell 6 days after induction. Arrow, pronounced after hyperpolarization potentials (AHP). **e**, Representative traces of action potentials in response to step current injections 15 days after induction. Membrane potential was maintained at approximately -63 mV. **f**, Quantification of intrinsic membrane properties in control ES cells (0 day) before and after viral transduction. Rin, membrane input resistance; RMP, resting membrane potential; Cm, capacitance. Scale bars, 10 μ m (**a, b, c**). Numbers of cells recorded are labelled in the bars. Note the heterogeneity of the parameters (see also Supplementary Fig. 1). Data are presented with mean \pm s.e.m. ***P < 0.05.

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ²Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ³Program in Cancer Biology, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ⁴Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA. ⁵Howard Hughes Medical Institute, Stanford University School of Medicine, 265 Campus Drive, Stanford, California 94305, USA.

*These authors contributed equally to this work.

after-hyperpolarization potentials (AHPs) following action potentials (Fig. 1d, f). Similar findings could be observed when human induced pluripotent stem (iPS) cells were infected (Supplementary Fig. 3). Thus, the BAM factors rapidly induce neuronal differentiation of human pluripotent stem cells.

Next, we asked whether human fibroblasts could also be directly converted into neurons. To this end, we derived three independent primary human fetal fibroblast lines (HFFs) (see Methods) and performed an extensive characterization of these cultures in various growth conditions to confirm that they lack spontaneous neuronal differentiation potential and do not contain detectable amounts of neural crest stem cells (see Supplementary Fig. 4). Strikingly, 7–10 days after infection with the BAM factors we could detect cells with immature neuronal morphologies. These cells expressed Tuj1 (Supplementary Fig. 5a), but remained functionally immature as revealed by their inability to generate action potentials 20 days after doxycycline treatment (Supplementary Fig. 5b). Thus, the BAM factors seemed to induce neuronal features but were insufficient to generate functional neurons from human fetal fibroblasts under these conditions.

Therefore, we screened 20 additional factors that could improve the generation of neuronal cells in combination with the BAM pool. We observed that *NeuroD1*, another basic helix–loop–helix transcription factor, improved the efficiency of generating Tuj1-positive neuronal cells two to threefold after 3 weeks (Fig. 2a). To determine the relative contribution of *NeuroD1*, we tested various combinations of these four factors. *NeuroD1* alone had no effect, but surprisingly in combination with *Brn2* it was sufficient to generate a similar number of Tuj1-positive neuronal cells compared to the BAN (*Brn2*, *Ascl1* and *NeuroD1*), BMN (*Brn2*, *Myt1l* and *NeuroD1*) and BAMN (*Brn2*, *Ascl1*, *Myt1l* and *NeuroD1*) pools (Supplementary Fig. 6a). However, further morphological and functional characterization showed that the BAMN combination generated the most mature neuronal cells (Supplementary Fig. 6b). We therefore decided to focus the further analysis on BAMN-iN cells.

Two weeks after induction, BAMN-iN cells showed neuronal morphologies and were labelled with pan-neuronal antibodies such as anti-Tuj1, anti-NeuN (also known as Rbfox3), anti-PSA-NCAM (polysialylated neural cell adhesion molecule) and anti-MAP2 (Fig. 2b–f). After extended culture periods of 4–5 weeks, we could detect cells expressing neurofilaments (Supplementary Fig. 7a), and rare neuronal processes decorated with punctate staining of synapsin and synaptotagmin, two synaptic vesicle proteins (Fig. 2g, h and Supplementary Fig. 7b). To ensure the co-expression of pan-neuronal and subtype specific markers, we performed single-cell gene-expression profiling of iN cells using Fluidigm dynamic RT-PCR arrays¹⁸. We analysed 54 single HFF-iN cells 34 days after doxycycline treatment from two independently infected cultures (Supplementary Fig. 8). These data revealed robust co-expression of multiple pan-neuronal and synaptic markers in 50/54 HFF-iN cells (β -III-tubulin, *DCX*, *MAP2*, *NCAM*, synapsin). Over half (29 of 54) of the iN cells analysed expressed mRNAs typical for glutamatergic neurons, such as *vGLUT1*, *vGLUT2* (also known as *SLC17A7* and *SLC17A7*, respectively) or both (Fig. 2i). Only two iN cells expressed *GAD67* (also known as *Gad1*) in the absence of *vGLUT1* or 2, and no iN cell expressed the inhibitory marker gene *vGAT* (also known as *SLC32A1*). Five cells expressed the catecholaminergic neuron marker tyrosine hydroxylase. Immunofluorescence analysis revealed that 5 weeks after infection $17 \pm 8\%$ of iN cells expressed the forebrain marker *Tbr1*, $21 \pm 9\%$ expressed the marker of peripheral neurons peripherin, whereas *En1*, a marker of midbrain neurons, serotonin and choline acetyltransferase were not detectable (Supplementary Fig. 9).

To assess whether the iN cell state was stable without continued transgene expression, we monitored the mRNA expression levels of endogenous and exogenous BAMN genes before and after doxycycline induction and after doxycycline withdrawal. Whereas the exogenous transgenes were clearly doxycycline-dependent, the four corresponding endogenous genes were rapidly induced and exhibited increasing

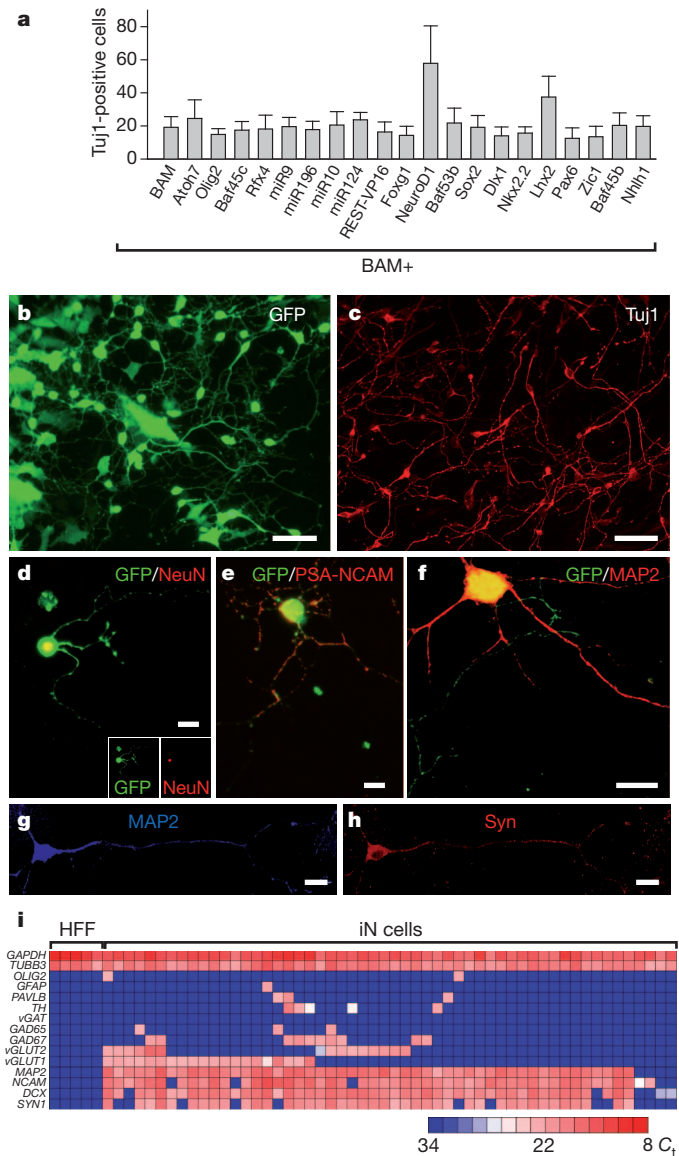


Figure 2 | NeuroD1 increases reprogramming efficiency in primary human fetal fibroblasts. **a**, Quantification of Tuj1-positive BAM-iN cells with indicated factors, 3 weeks after doxycycline treatment. **b**, **c**, Three weeks after doxycycline treatment BAM+NeuroD1 iN cells showed neuronal morphologies (**b**) and expressed Tuj1 (**c**). **d**–**f**, iN cells expressed NeuN (**d**), PSA-NCAM (**e**), and MAP2 (**f**) 2 weeks after doxycycline treatment. **g**, **h**, An iN cell expressing MAP2 (**g**) and synapsin (**h**) 4 weeks after doxycycline treatment and co-cultured with primary astrocytes. **i**, Single-cell gene expression profiling using Fluidigm dynamic arrays. Rows represent the evaluated genes and columns represent individual cells. Heat map (blue to red) represents the threshold C_t values as indicated. Data in (**a**) are presented as mean \pm s.d. Scale bars, 100 μ m (**b**, **c**) and 10 μ m (**d**–**h**).

expression levels over time even after doxycycline withdrawal (Supplementary Fig. 10). Similarly, HFF-iN cells could be maintained in the absence of doxycycline for 3 weeks (Supplementary Fig. 11).

We next asked whether iN cells generated from HFFs had active membrane properties. iN cells were identified by EGFP fluorescence (Fig. 3b) and whole-cell recordings were performed 14–35 days after doxycycline treatment. Two to three weeks after addition of doxycycline the average resting membrane potential of HFF-iN cells was -52.2 ± 2.2 mV (mean \pm s.e.m., $n = 41$). When HFF-iN cells were step-depolarized, action potentials could be detected in many iN cells at 14–25 days, and in all recorded iN cells at days 34–35 (Fig. 3c, Supplementary Figs 7 and 12 and Supplementary Table 1). Fast-activating and inactivating inward Na^+ currents as well as outward K^+ currents were also observed (Fig. 3d).

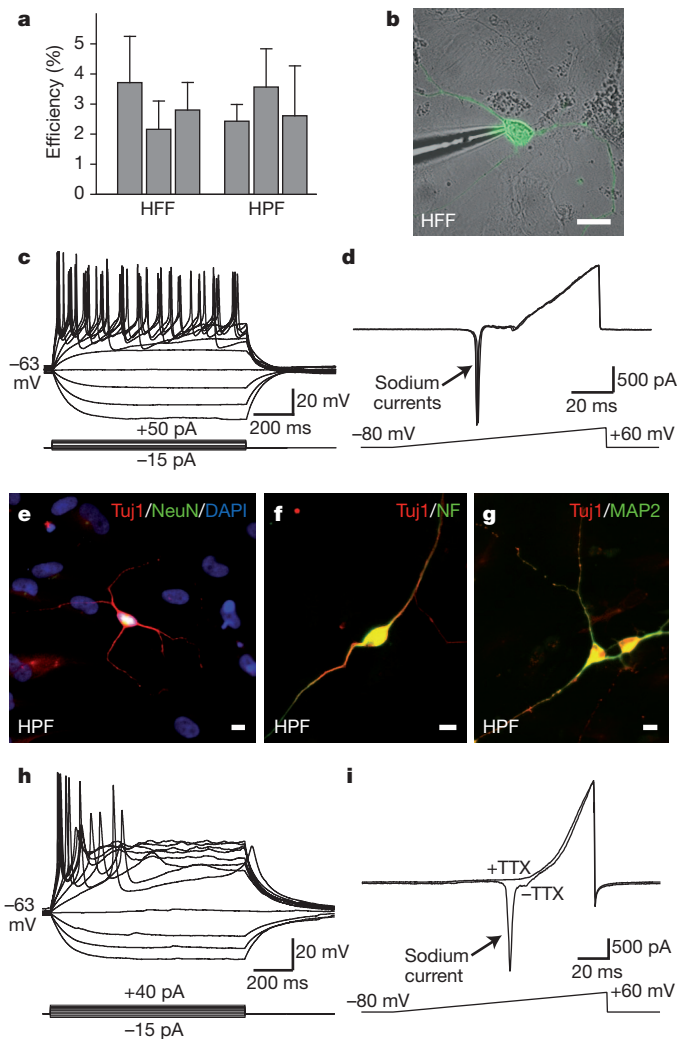


Figure 3 | Membrane properties of fibroblast iN cells. **a**, Quantification of Tuj1-positive neuronal cells from HFFs (line HFF-A) 3 weeks after doxycycline treatment or HPFs (line HPF-B) 4 weeks after doxycycline treatment. $n = 3$ independent experiments. **b**, Patch clamp recording was conducted on HFF-iN cells identified by EGFP fluorescence and differential interference contrast microscopy. **c**, Representative traces of membrane potentials in response to step current injections (lower panel) from an HFF-iN cell 19 days after doxycycline treatment. Membrane potential was maintained at approximately -63 mV. **d**, Representative traces of membrane currents recorded with a ramp protocol (lower panel). Fast activating and inactivating Na^+ currents were prominent. Three traces are shown superimposed. **e–g**, HPF-iN cells express Tuj1 (red) and NeuN (green) (**e**), neurofilament (green, NF) (**f**) and MAP2 (green) (**g**). **h**, Representative traces of membrane potentials in response to step current injections in HPF-iN cells. Action potentials were generated in cultures without glia. **i**, Representative traces of membrane currents recorded following a ramp protocol (lower panel) in HPF-iN cells. The Na^+ currents could be blocked by tetrodotoxin (TTX). Data in **a** are presented as mean \pm s.d. Scale bars, $10\ \mu\text{m}$ (**a**, **e–g**).

To determine whether the BAMN factors were also capable of converting more mature human fibroblasts into iN cells, we derived primary human postnatal fibroblasts (HPFs) from three different perinatal foreskin resections. In all three HPF lines, expression of BAMN factors reproducibly generated neuron-like cells with co-expression of multiple pan-neuronal markers (Fig. 3e–g). Intriguingly, the efficiencies of iN cell generation from fetal and postnatal fibroblasts were similar (2–4% of cells plated; Fig. 3a). Single-cell gene-expression profiling of iN cells revealed that 46 of 51 HPF iN cells co-expressed pan-neuronal and synaptic markers 42 days after infection; the majority of the HPF iN cells (37 of 51) seem to be glutamatergic neurons. Immunofluorescence analysis showed that 6 weeks after infection, $81 \pm 17\%$ of iN cells

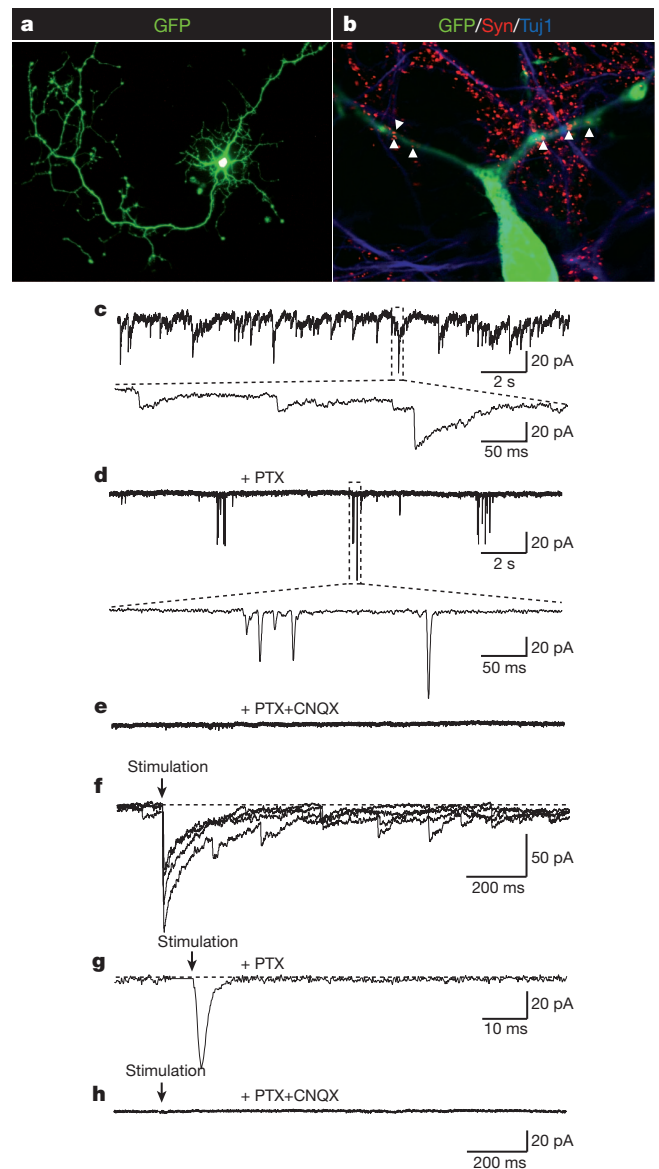


Figure 4 | Synaptic responses of HFF-iN cells. **a**, An HFF-iN cell expressing EGFP co-cultured with mouse cortical neurons at day 35 after doxycycline treatment. **b**, Synapsin-positive puncta co-localize with neurites extending from HFF-iN cells (arrow heads). **c**, Thirty-five days after doxycycline treatment, spontaneous PSCs were recorded in HFF-iN cells. **d**, Slow responses could be blocked by picrotoxin (PTX). The insert shows the fast kinetics of the responses. **e**, In the presence of PTX and CNQX (both $50\ \mu\text{M}$), no spontaneous activities were observed. **f**, Evoked postsynaptic responses. Four traces were superimposed. **g**, In the presence of PTX, electric stimulation evoked fast-kinetic excitatory PSCs (EPSCs). **h**, No evoked synaptic responses were observed in the presence of PTX and CNQX. Scale bars, $100\ \mu\text{m}$ (**a**) and $10\ \mu\text{m}$ (**b**).

expressed Tbr1 above the levels of fibroblasts, and $15.2 \pm 6.6\%$ were peripherin-positive (Supplementary Fig. 13). Unlike fetal fibroblasts, most postnatal fibroblasts showed weak but specific Tbr1 staining (Supplementary Fig. 13a–d). Electrophysiological recordings demonstrated the presence of regenerative action potentials as well as voltage-dependent channel activities in the majority of cells analysed from two lines (for example, 17 of 18 cells from line HPF-B, Fig. 3h, i and Supplementary Table 1). Furthermore, iN cells with active membrane properties could be generated from dermal fibroblasts derived from an 11-year-old human subject (Supplementary Fig. 14).

Finally, we determined whether human iN cells can express functional neurotransmitter receptors and form functional synapses.

Application of either GABA (γ -aminobutyric acid) or L-glutamate to HFF-iN cells induced current responses that could be blocked by picrotoxin and CNQX, respectively (Supplementary Fig. 12e, f). We then dissociated HFFs 4–7 days after infection with the BAMN factors and EGFP and plated them onto previously established mouse cortical neuronal cultures. These co-cultures were maintained for up to 5 weeks thereafter. HFF-iN cells were identified by the EGFP expression (Fig. 4a). Whole-cell recordings after 2–3 weeks of co-culture showed no synaptic activity ($n = 20$) but after 4–5 weeks approximately half of human iN cells recorded showed spontaneous post-synaptic currents (PSCs) with variable kinetics ($n = 21$, Fig. 4c and Supplementary Table 1). Immunostaining with synapsin antibodies confirmed the presence of scattered synaptic puncta on the dendrites of EGFP-positive cells (Fig. 4b). When the GABA_A receptor inhibitor picrotoxin was applied, the majority of the spontaneous PSCs were blocked, demonstrating that they were inhibitory (IPSCs) (Fig. 4d). In the presence of picrotoxin, bursting of spontaneous excitatory post-synaptic currents (EPSCs) was shown and could be blocked by the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor blocker CNQX (Fig. 4e). Focal stimulation evoked both IPSCs and EPSCs that could be blocked by picrotoxin and CNQX (Fig. 4f–h). Importantly, PSCs could also be recorded from HPF-iN cells co-cultured with mouse cortical neurons 4 weeks after infection (Supplementary Fig. 15). These results demonstrate that fetal and postnatal fibroblast-derived iN cells could form functional synapses and integrated into pre-existing neuronal networks.

In this report, we have identified a combination of transcription factors that are capable of converting human fibroblasts directly into neurons. Like mouse iN cells¹³ and neurons derived from ES cells^{19–21} and iPS cells^{22,23}, the human iN cells seem relatively immature, as indicated by their slightly depolarized membrane potentials and the relatively low-amplitude synaptic responses. Compared to mouse iN cells, human iN cells required longer culture periods to develop synaptic activity. Future studies will be necessary to thoroughly optimize conditions for human iN cell generation and maturation, which would facilitate applications of this method for the study of human neuronal development and disease.

METHODS SUMMARY

Cell culture. H9 human ES cells (WiCell Research Resources) and iPS cells were expanded in mTeSR1 (Stem Cell Technologies) and passaged as clumps or single cells²⁴. Primary HFFs were isolated from the distal half of the limbs of GW8-10 fetuses obtained from Advanced Bioscience Resources Inc. Primary HPFs were established from foreskin. Primary mouse cortical cultures and glial monolayer cultures were established as described previously¹³.

Lentiviral infections. Lentiviral production and fibroblast infections were performed as described previously¹³. Primary fibroblasts and pluripotent stem cells were infected with concentrated lentivirus and treated with doxycycline ($2 \mu\text{g} \mu\text{l}^{-1}$) 16–24 h later.

Electrophysiology and expression analysis. Cells were analysed by immunofluorescence and electrophysiology as described elsewhere^{13,25}. Single-cell gene expression profiling was performed using the Fluidigm Biomark dynamic array^{18,26} according to the manufacturer's instructions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 November 2010; accepted 18 May 2011.

Published online 26 May 2011.

- Blau, H. M. *et al.* Plasticity of the differentiated state. *Science* **230**, 758–766 (1985).
- Gurdon, J. B. From nuclear transfer to nuclear reprogramming: the reversal of cell differentiation. *Annu. Rev. Cell Dev. Biol.* **22**, 1–22 (2006).
- Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neurosci.* **5**, 308–315 (2002).
- Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).

- Shen, C.-N., Slack, J. M. W. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* **11**, 1553–1558 (2001).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to β -cells. *Nature* **455**, 627–632 (2008).
- Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
- Zhou, Q. & Melton, D. A. Extreme makeover: converting one cell into another. *Cell Stem Cell* **3**, 382–388 (2008).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Kriegstein, A., Noctor, S. & Martinez-Cerdeno, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7**, 883–890 (2006).
- Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
- Bottenstein, J. E. & Sato, G. H. Growth of a rat neuroblastoma cell line in serum-free supplemented medium. *Proc. Natl Acad. Sci. USA* **76**, 514–517 (1979).
- Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
- Johnson, M. A., Weick, J. P., Pearce, R. A. & Zhang, S. C. Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. *J. Neurosci.* **27**, 3069–3077 (2007).
- Wu, H. *et al.* Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. *Proc. Natl Acad. Sci. USA* **104**, 13821–13826 (2007).
- Koch, P., Opitz, T., Steinbeck, J. A., Ladewig, J. & Brustle, O. A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for *in vitro* instruction and synaptic integration. *Proc. Natl Acad. Sci. USA* **106**, 3225–3230 (2009).
- Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- Hu, B. Y. *et al.* Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc. Natl Acad. Sci. USA* **107**, 4335–4340 (2010).
- Xu, Y. *et al.* Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc. Natl Acad. Sci. USA* **107**, 8129–8134 (2010).
- Maximov, A., Pang, Z. P., Tervo, D. G. & Sudhof, T. C. Monitoring synaptic transmission in primary neuronal cultures using local extracellular stimulation. *J. Neurosci. Methods* **161**, 75–87 (2007).
- Wong, C. C. *et al.* Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnol.* **28**, 1115–1121 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank Y. Kokubu for technical assistance in molecular cloning and Y. Zhang for assistance in iPS cell induced neuron culture. We also thank Y. Sun for providing the microRNAs expression lentiviral vectors and S. Majumder for the REST-VP16 construct. This work was enabled by start-up funds from the Institute for Stem Cell Biology and Regenerative Medicine at Stanford (M.W.), the Ellison Medical Foundation (M.W.), the Stinehard-Reed Foundation (M.W.), the Donald E. and Delia B. Baxter Foundation (M.W.), the NIH grants 1R01MH092931 (M.W. and T.C.S.) and RC4 NS073015 (M.W.), and a Robertson Investigator Award from the New York Stem Cell Foundation. Z.P.P. is supported by 2008 and 2010 NARSAD Young Investigator Awards. T.V. is supported by the Ruth and Robert Halperin Stanford Graduate Fellowship. A.C. is supported by the AXA research fund and D.R.F. is supported by BioX Undergraduate Fellowship.

Author Contributions Z.P.P., N.Y., T.V., A.O., T.C.S. and M.W. designed the experiments and analysed the data. D.R.F. and T.Q.Y. helped with lentiviral production. A.C., V.S. and S.M. helped to provide experimental material and helped with the analyses. Z.P.P., N.Y., T.V., T.C.S. and M.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.W. (wernig@stanford.edu).

METHODS

Cell culture. H9 human ES cells (WiCell Research Resources) were expanded in mTeSR1 (Stem Cell Technologies). Induced pluripotent stem cells were generated as described elsewhere²⁷. The day before infection, cells were treated with Accutase and seeded as single cells in 3.5-cm tissue culture dishes on Matrigel in mTeSR1 containing 2 μ M thiazovivin (Bio Vision)²⁴. To inhibit the growth of uninfected ES cells and select for post-mitotic neurons, we added 4 μ M cytosine β -D-arabinofuranoside (Ara-C) to the media 48 h after doxycycline addition. Primary human fetal fibroblasts were isolated from the distal half of the limbs of 8–10-week-old fetuses (Advanced Bioscience Resources Inc.). The tissue was dissociated after trypsin digestion and plated in MEF media (DMEM high glucose, calf serum, sodium pyruvate, non-essential amino acids, penicillin/streptomycin and β -mercaptoethanol). Primary human postnatal fibroblasts (HPFs) were established from dissociated foreskin tissue derived from 1–3-day-old newborns. Before being used for experiments, primary fibroblast cells were passaged at least three times. Primary mouse cortical cultures and glial monolayer cultures were established as described previously¹³. To maintain the iN cell culture, cells were either grown in N3 medium (DMEM/F2 (Invitrogen), apotransferrin (100 μ g ml⁻¹), insulin (5 μ g ml⁻¹), sodium selenite (30 nM), progesterone (20 nM), putrescine (100 nM), penicillin/streptomycin) supplemented with neurotrophic factors including brain-derived neurotrophic factor, glial-cell-derived neurotrophic factor, neurotrophin-3 and ciliary neurotrophic factor (R&D systems), or dissociated using papain and replated onto previously established monolayer culture of primary mouse glia or primary neurons from mouse cortex in neuronal growth medium (MEM (Gibco) supplemented with B27 (Gibco), glucose (5 mg ml⁻¹), transferring (10 μ g ml⁻¹), 5% fetal bovine serum and Ara-C (2 μ M, Sigma)^{13,25}.

Virus infections. Lentiviral production and fibroblast infections were performed as described previously¹³. Briefly, primary human fetal or postnatal fibroblasts were plated and infected with concentrated lentiviral particles and polybrene (8 μ g μ l⁻¹) in fresh MEF medium. Viral medium was removed after 16–24 h and replaced with MEF medium containing doxycycline (2 μ g μ l⁻¹). After 24–48 h, medium was changed to N3 medium containing doxycycline (2 μ g μ l⁻¹). For human ES cell infections, H9 human embryonic stem cells were switched into N3 medium containing polybrene (2 μ g μ l⁻¹) 24 h after re-plating, and concentrated lentiviral particles were added. After 16–24 h, cultures were switched to N3 medium containing doxycycline (2 μ g μ l⁻¹) and changed daily before dissociation. Forty-eight hours after the initial addition of doxycycline, Ara-C (4 μ g μ l⁻¹) was added to the medium to inhibit proliferation of uninfected ES cells until analysis 6 days after the addition of doxycycline. All chemicals were purchased from Sigma-Aldrich if not otherwise specified.

Single-cell gene expression analysis (Fluidigm dynamic array). Single-cell gene expression profiling was performed using the Fluidigm Biomark dynamic array according to the manufacturer's protocol^{18,26}. Briefly, single cells growing on culture dishes after 5 or 6 weeks of transduction were collected by aspiration into patch electrodes and ejected into 2 \times cellsdirect buffer (Invitrogen), flash-frozen and kept at -80°C until further processing. Thawed cells were subject to target-specific reverse-transcription and 18 cycles of PCR pre-amplification with a mix of primers specific to the target genes (STA). STA products were then processed for real-time PCR analysis on Biomark 48:48 Dynamic Array integrated fluidic circuits (Fluidigm). To ensure the specificity of the amplification, titrations of total human brain RNA were included in each experiment, and only primers that demonstrated a linear amplification were analysed. Furthermore, melting curves of the PCR products were compared between the single cells and the control RNA to ensure the specificity of the PCR products.

Electrophysiology. Action potentials were recorded with current-clamp whole-cell configuration. The pipette solution for current-clamp experiments contained (in mM): 123 K-gluconate, 10 KCl, 1 MgCl₂, 10 HEPES, 1 EGTA, 0.1 CaCl₂, 1 K₂ATP, 0.2 Na₄GTP and 4 glucose, pH adjusted to 7.2 with KOH. Membrane potentials were kept around -65 to -70 mV, and step currents were injected to elicit action potentials. For whole-cell voltage-dependent current recordings, the same internal solution as aforementioned was used. For synaptic functional evaluation, the internal solution contained (in mM): 135 CsCl, 10 HEPES, 1 EGTA, 4 Mg-ATP, 0.4 Na₄GTP, and 10 QX-314, pH 7.4. The bath solution contained (in mM): 140 NaCl, 5 KCl, 2 CaCl₂, 2 MgCl₂, 10 HEPES, and 10 glucose, pH 7.4. Synaptic responses were measured as described previously^{13,25}. Stimulus artefacts for evoked synaptic responses were removed for graphic representation. Electrophysiological data are presented as mean \pm s.e.m. All statistical comparisons were made using Student's *t*-test.

Immunofluorescence and RT-PCR. For immunofluorescence experiments, cells were fixed in 4% paraformaldehyde in PBS for 10 min at room temperature. After fixation, cells were incubated in 0.2% Triton X-100 in PBS for 5 min at room temperature. After washing twice with PBS, cells were blocked in a solution of PBS containing 4% BSA and 1% Cosmic calf serum (CCS) for 30 min at room temperature. Primary and secondary antibodies were applied for 1 h and 30 min, respectively. The following antibodies were used for our analysis: rabbit anti-Tuj1 (Covance, 1:1,000), mouse anti-Tuj1 (Covance, 1:1,000), mouse anti-MAP2 (Sigma, 1:500), mouse anti-NeuN (Millipore, 1:200), mouse anti-neurofilament (Developmental Studies Hybridoma Bank (DSHB), 2H3a, 1:1,000), rabbit anti-synapsin (E028, 1:1,000), anti-synaptotagmin (Synaptic systems, 41.1, 1:2,000), guinea pig anti-vGLUT1 (Millipore, 1:2,000), mouse anti-GAD6 (DSHB, 1:500), rabbit anti-Tbr1 (Abcam, 1:200), mouse anti-peripherin (Sigma, 1:100), sheep anti-tyrosine hydroxylase (Pel-Freez, 1:500), rabbit anti-GFAP (DAKO, 1:4,000), mouse anti-Sox2 (R&D Systems, 1:50), goat anti-Brn2 (clone C-20, Santa Cruz Biotechnology, 1:100), rabbit anti-Ascl1 (Abcam, 1:200), mouse anti-BrdU (Becton Dickinson, 1:50), mouse anti-LU5 (Abcam, 1:200), goat anti-Sox10 (Santa Cruz Biotechnology, 1:40). Fluorescein isothiocyanate (FITC)-, and Cy3-conjugated secondary antibodies were obtained from Jackson ImmunoResearch. Alexa-488-, Alexa-546- and Alexa-633-conjugated secondary antibodies were obtained from Invitrogen. 4',6-Diamidino-2-phenylindole (DAPI) was from Sigma (1:10,000). For RT-PCR analysis, RNA was isolated using an RNAqueous Kit (Applied Biosystems) following the manufacturer's instructions, treated with DNase (Applied Biosystems) and reverse-transcribed with Superscript III (Invitrogen).

Efficiency calculation. The following method was used to calculate the efficiency of neuronal induction. The total number of Tuj1-positive cells with a neuronal morphology, defined as cells having a circular, three-dimensional appearance that extend a thin process at least three times longer than their cell body, were quantified at indicated time points. We determined this number in at least 15 randomly selected $\times 20$ visual fields on a Zeiss Axio Observer microscope. The average and standard deviation per field was determined and then used to extrapolate the total number of iN cells present in the entire dish based the known surface areas of a $\times 20$ visual field and the respective culture dish. We then divided this number by the number of cells plated before infection to get the percentage of the starting population of cells that adopted neuron-like characteristics. In multiple independent experiments we verified that this extrapolation method yields cell numbers very similar to those measured by a haemocytometer at the time of plating. Data are presented as mean \pm s.d.

27. Somers, A. *et al.* Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells* **28**, 1728–1740 (2010).

Direct generation of functional dopaminergic neurons from mouse and human fibroblasts

Massimiliano Caiazzo¹, Maria Teresa Dell'Anno^{1*}, Elena Dvoretzskova^{2*}, Dejan Lazarevic^{3,4}, Stefano Taverna², Damiana Leo², Tatyana D. Sotnikova², Andrea Menegon⁵, Paola Roncaglia⁴, Giorgia Colciago¹, Giovanni Russo², Piero Carninci⁶, Gianni Pezzoli⁷, Raul R. Gainetdinov², Stefano Guscini^{4,8}, Alexander Dityatev² & Vania Broccoli¹

Transplantation of dopaminergic neurons can potentially improve the clinical outcome of Parkinson's disease, a neurological disorder resulting from degeneration of mesencephalic dopaminergic neurons^{1,2}. In particular, transplantation of embryonic-stem-cell-derived dopaminergic neurons has been shown to be efficient in restoring motor symptoms in conditions of dopamine deficiency^{3,4}. However, the use of pluripotent-derived cells might lead to the development of tumours if not properly controlled⁵. Here we identified a minimal set of three transcription factors—*Mash1* (also known as *Ascl1*), *Nurr1* (also known as *Nr4a2*) and *Lmx1a*—that are able to generate directly functional dopaminergic neurons from mouse and human fibroblasts without reverting to a progenitor cell stage. Induced dopaminergic (iDA) cells release dopamine and show spontaneous electrical activity organized in regular spikes consistent with the pacemaker activity featured by brain dopaminergic neurons. The three factors were able to elicit dopaminergic neuronal conversion in prenatal and adult fibroblasts from healthy donors and Parkinson's disease patients. Direct generation of iDA cells from somatic cells might have significant implications for understanding critical processes for neuronal development, *in vitro* disease modelling and cell replacement therapies.

Seminal studies have demonstrated that functional neurons can be generated independently of stem cells by direct cell conversion through genetics-based approaches⁶. More recently, in a set of elegant experiments, fibroblasts have been directly converted into neuronal cells (iNs) by the forced expression of the three neurodevelopmental factors *Mash1*, *Brn2* (also known as *Pou3f2*) and *Myt1l*⁷. However, iNs represent a heterogeneous population of glutamatergic and GABAergic neurons and their degree of global reprogramming remains to be properly characterized. It is thus unclear whether a specific neuronal subtype can be preferentially induced from direct reprogramming of heterologous cells. Therefore, we aimed to generate dopaminergic neurons through the direct conversion of somatic cells by forced expression of lineage-specific factors that act during brain development^{8,9}. Initially, we transduced mouse embryonic fibroblasts (MEFs) from TH-GFP transgenic mice¹⁰ with a mixture of doxycycline (dox)-inducible lentiviruses expressing all selected factors (11 dopaminergic and 3iN; Supplementary Table 1) or with DsRed retrovirus (negative control) (Fig. 1a–d). We did not observe any GFP⁺ cells in MEFs 10 days after Ds-Red retrovirus infection or in culture without any viral infection (Fig. 1d, inset). In contrast, transduction of all factors resulted in the generation of a small number of bright GFP⁺ cells ($1.8 \pm 0.8\%$) (Supplementary Fig. 1d–f). We next sought to determine the minimal set of genes required for dopaminergic neuronal induction. Given its essential role as a proneural gene during neurogenesis, *Mash1* was

introduced into MEFs together with each other single dopaminergic factor. Reporter gene expression was elicited only when *Mash1* was combined with *Nurr1*, a critical determinant of dopaminergic neuronal specification and survival during development and in adulthood¹¹. However, *Mash1/Nurr1* combined activation elicited a modest increase of GFP⁺ cells ($8 \pm 2\%$) (Supplementary Fig. 1g–i). Therefore, we added a third molecule of the 12 remaining and scored for the rate and morphology of GFP⁺ cells in each combination. Surprisingly, only *Lmx1a* and in part *Lmx1b* ($18 \pm 3\%$ versus $13 \pm 3\%$ of GFP⁺ cells, respectively) were able to synergize with *Mash1/Nurr1*, robustly increasing the generation of GFP⁺ cells with an evidently complex neuronal morphology (Fig. 1h and Supplementary Fig. 1s–y). Using the *Mash1*/

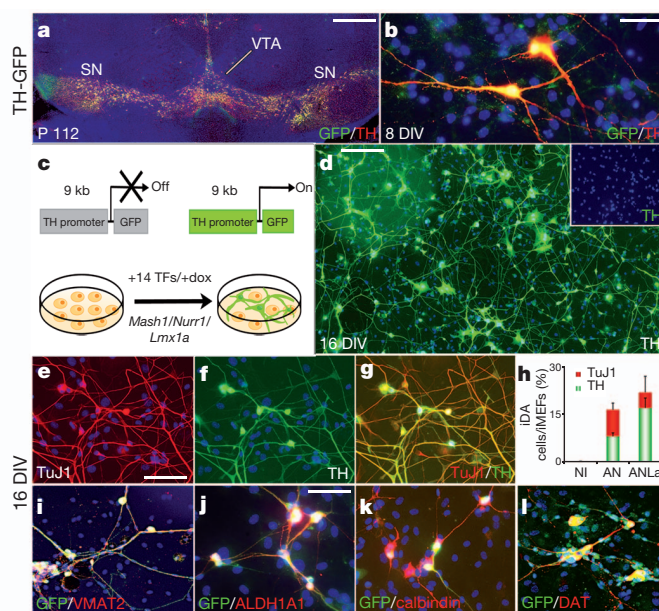


Figure 1 | *Mash1*, *Nurr1* and *Lmx1a* reprogram mouse fibroblasts into iDA cells. **a**, **b**, TH and GFP detection in TH-GFP adult brain (**a**) and ventral midbrain primary cell culture (**b**). DIV, days *in vitro*. SN, substantia nigra; VTA, ventral tegmental area. P112, postnatal day 112. **c**, Scheme of dopaminergic transcription factors (TFs) screening. **d**, TH staining in iDA cells and in uninfected MEFs (inset) after 16 days *in vitro*. **e–g**, iDA cells are positive for the dopaminergic markers TH (**e–g**), VMAT2, ALDH1A1, calbindin and DAT (**i–l**). **h**, Quantification of TuJ1⁺ and TH⁺ cells. AN, *Mash1*, *Nurr1*; ANLa, *Mash1*, *Nurr1*, *Lmx1a*; NI, non-induced. Data are presented as mean \pm s.e.m. Scale bars: 500 μ m (**a**), 20 μ m (**b**, **j**), 50 μ m (**e–g**, **k**, **l**) and 100 μ m (**d**, **i**).

¹Stem Cells and Neurogenesis Unit, Division of Neuroscience, San Raffaele Scientific Institute, 20132 Milan, Italy. ²Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genoa, Italy. ³CBM Srl. Area Science park, Basovizza, SS14, km165, 34149 Trieste, Italy. ⁴Sector of Neurobiology, International School for Advanced Studies (SISSA), via Bonomea, 265, 34136 Trieste, Italy. ⁵Advanced Light and Electron Microscopy Bio-Imaging Centre, Experimental Imaging Centre, San Raffaele Scientific Institute, 20132 Milan, Italy. ⁶Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-chô, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ⁷Parkinson Institute, Istituti Clinici di Perfezionamento, 20126 Milan, Italy. ⁸The Giovanni Armenise-Harvard Foundation Laboratory, 34136 Trieste, Italy.

*These authors contributed equally to this work.

Nurr1/Lmx1a factor combination, double GFP⁺/TH⁺ cells represented the majority of the induced TuJ1 neuronal cells (85 ± 4%). Supplementation of a fourth factor among the remaining ones failed to produce any further increase in GFP⁺ cells, with *Brn2* and *Myt1l*, the other two iN factors, even reducing the overall reprogramming efficiency (data not shown). For these reasons we focused on cells reprogrammed exclusively with the *Mash1/Nurr1/Lmx1a* factor combination. The same gene cocktail was also proficient in reprogramming adult mouse fibroblasts with high efficiency (Supplementary Fig. 2).

Sixteen days after reprogramming, a large number of GFP⁺ cells expressed many of the distinctive components of the dopaminergic machinery, such as TH, vesicular monoamine transporter 2 (VMAT2; also known as SLC18A2), dopamine transporter (DAT; also known as SLC6A3), as well as aldehyde dehydrogenase 1a1 (ALDH1A1) and calbindin (Fig. 1e–l). Conversely, markers associated with adrenergic (dopamine-β-hydroxylase (DBH)) or serotonergic (tryptophan hydroxylase 1 or 2 (TPH1/2)) serotonin transporter (SERT; also known as SLC6A4) neurons were not induced (data not shown). Transcriptional analysis by reverse transcription–polymerase chain reaction (RT–PCR) confirmed the activation of the dopamine-specific gene network including the endogenous expression of *Nurr1* and *Lmx1a* (Supplementary Fig. 3). Global expression analysis showed that iDA cells clustered with A9 and A10 adult mesencephalic dopaminergic (mDA) neurons rather than with their fibroblasts of origin as illustrated by hierarchical clustering (Fig. 2a, b) and the general degree of gene expression overlap (Fig. 2c). Of note, many representative genes of the dopaminergic phenotype, such as *Th*, *Vmat2*, *Aadc* (also known as *Ddc*), *Ret*, *Gfra1*, *Foxa1*, *Gdnf* and *Drd2* were highly enriched (Fig. 2d). Conversely, genes coding for adrenergic and serotonergic biosynthetic enzymes were found to be not upregulated in the reprogrammed cells (Fig. 2e). Moreover, the fibroblast markers *Twist2*, *Zeb2*, *Tgfb1i1* and *Chd2* (ref. 12) were downregulated in iDA cells (Fig. 2f). These findings indicate that the genetic reprogramming erased the majority of the evident expression hallmarks of the cell of origin, while specifically inducing the dopaminergic neuronal phenotype and not those of other

closely related neuronal subtypes. It should be noted that iDA expression profiling was close but distinguishable from that of mDA neurons, with 160 genes differently expressed with a ≥5-fold change (Supplementary Table 3). Whether this might indicate the presence of residual fibroblast gene expression in iDA cells remains to be addressed.

Interestingly, *Th* and *Vmat2* promoter regions were highly demethylated in dopaminergic neuronal cells whereas they were fully methylated in parental fibroblasts, indicating their epigenetic reactivation during dopaminergic neuronal conversion (Supplementary Fig. 4).

GFP⁺ cells induced by the three factors showed an elaborate neuronal morphology with multiple and long processes (Fig. 1d–l). Hence, we asked whether induced neuronal cells establish synaptic contacts in culture. Notably, synaptic resident proteins such as synaptotagmin I (SYT1) and synapsin (SYN) were localized in discrete puncta and colocalized with TH immunolabelling, suggesting the establishment of dopaminergic synaptic terminals (Supplementary Fig. 5). Moreover, successful FM4-64 dye uptake at TH⁺ synaptic boutons indicated active synaptic processes (Supplementary Fig. 5).

Next, we performed patch-clamp recordings of GFP⁺ iDA cells (*n* = 16) as well as primary mDA neurons (*n* = 12) to compare their respective physiological properties^{13,14}. iDA cells had higher cell resistance and lower capacitance than primary dopaminergic neurons, but showed normal resting membrane potential, normal Na⁺ currents (Fig. 3a), overshooting action potentials (Fig. 3c), and even more prominent K⁺ currents (Fig. 3b) and afterspike hyperpolarizations (Supplementary Table 4). More than 80% of iDA cells showed rhythmic discharges (Fig. 3d, e) at an average frequency of 2.6 Hz. The identity of voltage-gated inward Na⁺ and outward K⁺ currents in iDA cells has been verified pharmacologically (Supplementary Fig. 6). Next, we noted that iDA cells, like mDA neurons, express high levels of the D2 receptor (Fig. 3f). To verify whether the dopamine receptors were functional, we applied the specific D2/D3 receptor agonist quinpirole (1 μM), which markedly suppressed neuronal firing in 6 of 10 recorded cells in a reversible manner (Fig. 3g, h). Next, we used amperometry for real-time electrochemical detection of monoamine secretion from iDA cells^{15,16}. When carbon fibre electrodes were placed adjacent to GFP⁺ cells (Fig. 3i), depolarization of 4 cells with 25 mM K⁺ resulted in numerous amperometric events (Fig. 3i), reflecting quantal secretion of monoamines. Furthermore, high-performance liquid chromatography (HPLC) measurements revealed that iDA cells contain high levels of intracellular dopamine, detectable in pellet preparations, which can be released in the extracellular medium on stimulation with 50 mM KCl (Fig. 3j). Thus, reprogrammed cells show several major properties of dopaminergic neurons in terms of spontaneous spiking activity, temporal parameters of action potentials, inhibition of cell firing through D2 autoreceptors and controlled dopamine release.

Next, to determine the temporal requirements for the three exogenous factors to induce a stable reprogrammed cell state, infected MEFs were treated with dox for different time windows, after which dox was withdrawn. Only when fibroblasts were treated with dox for 6 or more days were numerous neuronal cells, mostly TH⁺, observed (Supplementary Fig. 7). Thus, reprogramming is a relatively rapid process that requires the expression of the three factors for only 6 days. At the same time, iDA cells achieved a stable neuronal state over time independently from viral transgene expression and even at 18 days of dox withdrawal, iDA cells were found at the same number, and exhibited similar spontaneous firing as iDA cells constantly cultivated in the presence of dox (Supplementary Fig. 8 and Supplementary Table 4).

Reprogramming of fibroblasts into differentiated neuronal cells might occur directly or by passing first through neural progenitors. When the DNA-base analogue BrdU was added from day 2 onwards to label the proliferating cells, virtually all neuronal cells were already post-mitotic after this time (Supplementary Fig. 9b, c, h). Despite the fact that during the first 2 days infected cells were actively proliferating in serum-containing medium, none showed expression of the neural progenitor molecular markers *Sox2*, *Ngn2*, *Otx2*, *Lmx1b* and

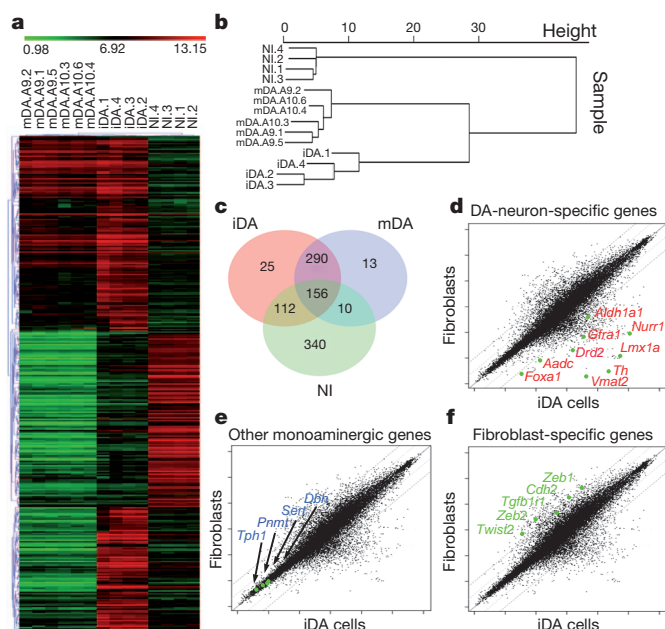


Figure 2 | Mouse iDA cells expression profiling. **a**, Heat-map of genes differentially expressed in RNA-microarray analysis performed on MEFs (NI), iDA cells and brain mDAs (A9–A10). **b**, **c**, Hierarchical clustering (**b**) and general degree of overlapping expression (**c**) among the three cell populations analysed. **d–f**, Scatter plots show that in iDA cells the majority of dopaminergic (DA) markers are increased, whereas other monoaminergic neuronal markers are not activated and fibroblast markers are silenced.

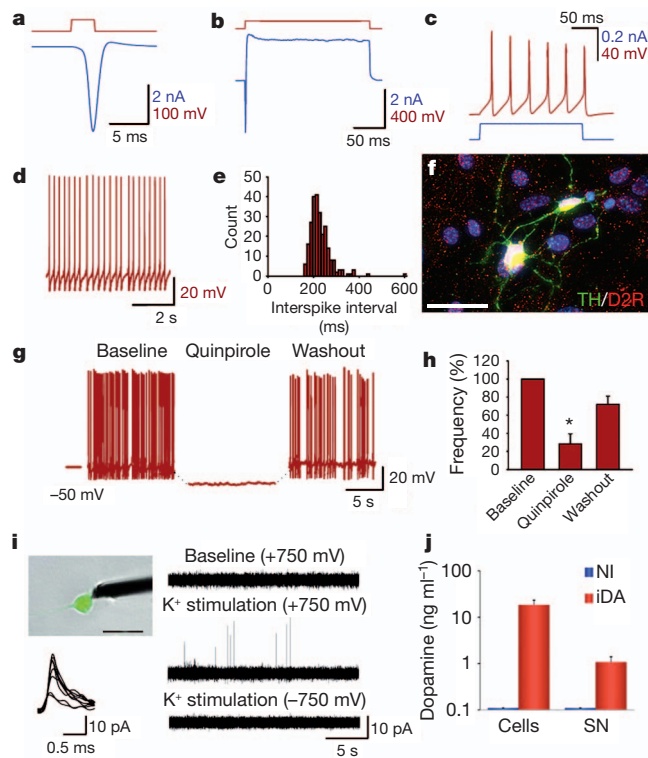


Figure 3 | Functional characterization of mouse iDA cells. **a, b,** Whole-cell voltage-clamp recording of Na⁺ and K⁺ currents. **c,** Current-clamp recording of multiple action potentials evoked by current injection. **d, e,** Current-clamp recording and interspike interval frequency of spontaneous action potentials. **f,** D2 receptor (D2R) staining. **g, h,** Effect of the D2/D3 agonist quinpirole on spiking frequency (**g**) and its statistical analysis (**h**) (**P* = 0.005, paired *t*-test, *n* = 6). **i,** Amperometric recordings of monoamine release after K⁺ stimulation; release events are shown with high-resolution below the image of the recorded cell. **j,** Dopamine content measured by HPLC in uninfected (NI) and iDA cells, both in cell pellets and in the supernatant (SN) after K⁺ stimulation. All cells are 16 days *in vitro* unless otherwise stated. Scale bars: 50 μm (**b**) and 20 μm (**i**). Data are presented as mean ± s.e.m.

En1 (ref. 17) (Supplementary Fig. 9i). Furthermore, we used a genetic tracing system based on activation of the Sox2^{β-geo} LacZ reporter¹⁸ showing that, as a proof-of-principle, LacZ activity was easily visualized upon reprogramming of Sox2^{β-geo} fibroblasts into induced pluripotent stem cells. By contrast, the reporter was never activated from the same cells when engaged into direct iDA reprogramming (Supplementary Fig. 9j). Altogether, these findings are inconsistent with the occurrence of detectable cell intermediates during the reprogramming of fibroblasts into iDA cells.

Next, the *in vivo* differentiation potential of iDA cells was assessed by orthotopic transplantations into neonatal mouse brains. Four days after viral transgene induction, infected cells were grafted into the ventricle of mouse newborn brains. Two and six weeks after transplantations, GFP⁺ cells were found integrated in the host tissue showing an extremely elaborated morphology (Supplementary Figs 10 and 11). Most of the GFP⁺ grafted neuronal cells were positive for TH, AADC, VMAT2 and DAT, indicating the achievement of a full neuronal dopaminergic cell fate (Supplementary Fig. 11b–g, i–l). Injection of brief supra-threshold current pulses evoked overshooting action potentials, and large Na⁺ and K⁺ currents were activated by depolarizing voltage steps (Supplementary Fig. 11o, p). Therefore, iDA cells maintain excitability and major currents *in vivo* even after an extensive period of time from grafting.

We then translated the same procedure to the human system by initially infecting IMR90 fetal fibroblasts. After 18 days from infection, we scored numerous TuJ1⁺ and TH⁺ neuronal cells accounting,

respectively, for 10 ± 4% and 6 ± 2% of the infected cells (Supplementary Fig. 12a–c). We then reprogrammed adult human fibroblasts from two healthy donors (aged 42 and 55) and from two patients with genetic forms of Parkinson's disease (Supplementary Table 5). Both healthy and diseased adult cells showed a comparable propensity to convert into neuronal cells, accounting for an estimated efficiency for TuJ1⁺ and TH⁺ cells of 5 ± 1% and 3 ± 1%, respectively (Fig. 4a–g). iDA cells were positive for ALDH1A1, TH, AADC, VMAT2 and DAT by immunocytochemistry (Fig. 4a–f) and gene expression analysis (Supplementary Fig. 13). Cell conversion was stable over time as the number and morphology of human iDA cells was not obviously affected up to day 24 after reprogramming even when dox was withdrawn from day 6 onwards (Fig. 4h, i).

Recordings in five infected fetal human iDA cells showed that the electrophysiological properties of these cells resemble mouse iDA cells (Supplementary Fig. 12). Recordings in eight infected adult human iDA cells revealed less mature phenotypes (Fig. 4j–l), with mean amplitudes: 0.7 ± 0.1 nA for Na⁺ currents; 1.2 ± 0.1 nA for delayed rectifier K⁺ currents; 78 ± 3 mV for action potentials; and 7.5 ± 2 mV for after-spike hyperpolarization. The identity of Na⁺ and K⁺ currents was confirmed pharmacologically (Fig. 4m, n). Most importantly, depolarization of

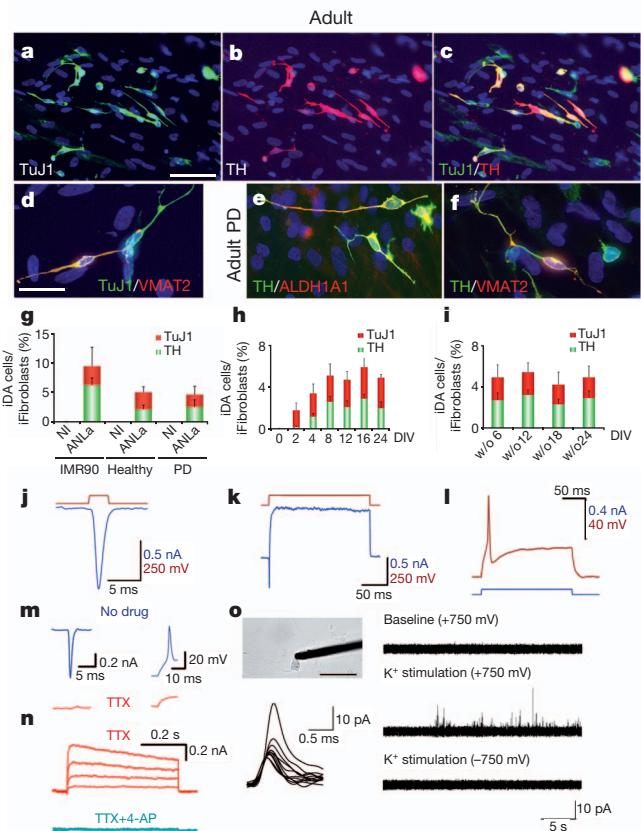


Figure 4 | Characterization of human fibroblasts reprogrammed into iDA cells. **a–f,** Fibroblasts from a healthy donor (**a–d**) and a Parkinson's disease (PD) patient (**e, f**) show a comparable efficiency in dopaminergic neuronal conversion. **g,** Quantification of iDA cells obtained from fetal (IMR90), healthy and Parkinson's disease adult fibroblasts. **h,** Quantification of TuJ1⁺ and TH⁺ cells in a time-course study from 0 to 24 days *in vitro* (DIV). **i,** Quantification of TuJ1⁺ and TH⁺ reprogrammed cells kept with (w) or without (w/o) doxycycline for 6, 12, 18 or 24 DIV. Data are presented as mean ± s.e.m. **j, k,** Whole-cell voltage-clamp recording of Na⁺ and K⁺ currents. **l,** Whole-cell current-clamp recording of single action potential elicited by a minimal depolarization. **m, n,** Suppression of Na⁺ (**m**) and K⁺ (**n**) currents and action potentials by tetrodotoxin (TTX) and 4-AP. **o,** Amperometric recordings of release events after K⁺ stimulation; high-resolution pattern is shown below the image of the recorded cell. All cells are 18 days *in vitro* unless otherwise stated. Scale bars: 50 μm (**a–c**) and 20 μm (**d–f, o**).

three cells with 25 mM K⁺ elicited numerous release events detected by amperometric measurements, as described earlier for mouse iDA cells (Fig. 4o). In summary, these experiments indicate that actively spiking dopamine-secreting cells can be induced by forced expression of the three factors in adult human cells from both healthy donors and Parkinson's disease patients.

Here, we demonstrated that the combination of three transcription factors—*Mash1*, *Nurr1* and *Lmx1a*—can rapidly and efficiently induce dopaminergic neuronal cells from mouse and human fibroblasts. Reprogrammed cells are similar to brain dopaminergic neurons in gene expression, and show dopamine release and pacemaker activity that can be modulated via D2 receptors. Importantly, this cell conversion diverges with respect to developmental neuronal lineage commitment as it does not progress through detectable intermediate neuronal stages. How this might impact on the overall functional aspects of the cells will need to be tested further in long-term *in vivo* transplantation studies.

Generation of functional dopaminergic neuronal cells by direct reprogramming opens new possibilities for regenerative therapies for Parkinson's disease and related disorders. However, to be clinically relevant, the overall cell conversion process needs to be highly efficient and the addition of supplementary factors might be helpful¹⁹. Interestingly, a different gene cocktail capable of inducing dopaminergic-like neurons from human fibroblasts has been identified²⁰. This opens the intriguing possibility that different molecular fate determinants reach a similar endpoint even though acting on different transcriptional cascades.

These procedures do not rely on pluripotent stem cells that are prone to tumours in their undifferentiated state. Moreover, the process described here does not pass through proliferative progenitors that also might be tumorigenic²¹. Thus, these procedures might avoid a dangerous drawback of stem cell therapies while providing a sufficient number of functional dopaminergic neurons amenable for autologous cell replacement therapies.

METHODS SUMMARY

Cell culture and viral infection. MEFs were isolated from embryonic day (E) 14.5 wild-type or TH-GFP mouse embryos. Adult human fibroblasts isolated from healthy subjects and Parkinson's disease patients as well as from human fetal lung fibroblasts (IMR90) were grown in MEF media. Cells were infected with dox-inducible lentiviruses as previously reported⁷.

Electrophysiology and amperometry. Electrophysiological recordings were performed in on-cell and whole-cell configurations. Carbon-fibre microelectrodes were used for amperometric recordings¹⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 March 2011; accepted 15 June 2011.

Published online 3 July 2011.

1. Lindvall, O. & Björklund, A. Cell therapy in Parkinson's disease. *NeuroRx* **1**, 382–393 (2004).
2. Politis, M. *et al.* Serotonergic neurons mediate dyskinesia side effects in Parkinson's patients with neural transplants. *Sci. Transl. Med.* **2**, 38–46 (2010).
3. Kim, J. H. *et al.* Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* **418**, 50–56 (2002).
4. Barberi, T. *et al.* Neural subtype specification of fertilization and nuclear transfer embryonic stem cells and application in parkinsonian mice. *Nature Biotechnol.* **21**, 1200–1207 (2003).
5. Wernig, M. *et al.* Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc. Natl Acad. Sci. USA* **105**, 5856–5861 (2008).

6. Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neurosci.* **5**, 308–315 (2002).
7. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
8. Ang, S.-L. Transcriptional control of midbrain dopaminergic neuron development. *Development* **133**, 3499–3506 (2006).
9. Smidt, M. P. & Burbach, J. P. How to make a mesodiencephalic dopaminergic neuron. *Nature Rev. Neurosci.* **8**, 21–32 (2007).
10. Sawamoto, K. *et al.* Visualization, direct isolation, and transplantation of midbrain dopaminergic neurons. *Proc. Natl Acad. Sci. USA* **98**, 6423–6428 (2001).
11. Perlmann, T. & Wallén-Mackenzie, A. Nurr1, an orphan nuclear receptor with essential functions in developing dopamine cells. *Cell Tissue Res.* **318**, 45–52 (2004).
12. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
13. Grace, A. A. & Bunney, B. S. The control of firing pattern in nigral dopamine neurons: single spike firing. *J. Neurosci.* **4**, 2866–2876 (1984).
14. Grace, A. A. & Onn, S. P. Morphology and electrophysiological properties of immunocytochemically identified rat dopamine neurons recorded *in vitro*. *J. Neurosci.* **9**, 3463–3481 (1989).
15. Pothos, E. N., Davila, V. & Sulzer, D. Presynaptic recording of quanta from midbrain dopamine neurons and modulation of the quantal size. *J. Neurosci.* **18**, 4106–4118 (1998).
16. Staal, R. G. W., Mosharov, E. V. & Sulzer, D. Dopamine neurons release transmitter via a flickering fusion pore. *Nature Neurosci.* **7**, 341–346 (2004).
17. Simeone, A. Genetic control of dopaminergic neuron differentiation. *Trends Neurosci.* **28**, 62–65 (2005).
18. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a β -geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).
19. Pang, Z. P. *et al.* Induction of human neuronal cells by defined transcription factors. *Nature* advance online publication, doi:10.1038/nature10202 (26 May 2011).
20. Pfisterer, U. *et al.* Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1105135108 (6 June 2011).
21. Amarglio, N. *et al.* Donor-derived brain tumor following neural stem cell transplantation in an ataxia telangiectasia patient. *PLoS Med.* **6**, e1000029 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to D. Bonanomi, S.-L. Ang, S. El Mestikawy, M. P. Smidt, M. German and F. Valtorta for providing valuable antibodies. We thank A. Sessa and V.B. laboratory members for helpful discussion. M. Wernig is acknowledged for providing the iN-inducing lentiviral vectors. We are thankful to S. Nicolis for sharing Sox2^{β-geo} mice. L. Muzio, C. Laterza and G. Martino are acknowledged for the generation of Sox2^{β-geo} induced pluripotent stem cells. M. Bacigaluppi is acknowledged for advice on stereological countings. We thank the “Cell Line and DNA Biobank” (G. Gaslini Institute) and “Human Genetic Bank of Patients affected by Parkinson Disease and parkinsonism” (Parkinson Institute of Milan) of the Telethon Genetic Biobank Network for human fibroblast samples. This study was supported by the “Fondazione Grigioni per il Morbo di Parkinson” (grant no. FGBRCVNI10310-001-V.B.), Eranet Neuron (V.B.), Cariplo Foundation (V.B.), Ministry of Health (Giovani ricercatori Award) (V.B.) and Italian Institute of Technology (V.B., A.D., S.G., T.S., R.G.).

Author Contributions M.C. and V.B. designed and conceived the experiments. M.C., M.T.D. and G.C. performed the lentiviral infections, characterized reprogrammed cells and analysed their fate after *in vivo* transplantation. E.D. and A.D. designed, performed and analysed all electrophysiological experiments. P.R., D.L., P.C. and S.G. performed the microarray gene expression profiling and analysed the data. D.L., A.D. and R.R.G. designed and D.L. and E.D. performed the amperometric experiments. R.R.G. and T.D.S. designed the protocol and performed the assessment of dopamine levels. S.T. and G.R. performed patch-clamp recording on brain slices. A.M. performed the functional analysis of synaptic activity. G.P. supervised the selection of the Parkinson's disease patients and the isolation of the primary fibroblasts. V.B. and A.D. should be considered as co-senior authors and wrote the manuscript.

Author Information Data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE27174 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27174>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.B. (broccoli.vania@hsr.it).

METHODS

Cell culture. MEFs were isolated from E14.5 wild-type or TH-GFP knock-in mice embryos. Head, vertebral column, dorsal root ganglia and all internal organs were removed and discarded and the remaining embryonic tissue was manually dissociated and incubated in 0.25% trypsin (Sigma) for 10–15 min. Cells from each embryo were plated onto a 15-cm tissue culture dish in MEF media (Dulbecco's modified Eagle medium (DMEM; Invitrogen) containing 10% fetal bovine serum (FBS; Hyclone), non-essential amino acids (Invitrogen), sodium pyruvate and penicillin/streptomycin (Invitrogen)). In all experiments cells were not split more than four times. Mouse adult fibroblasts were isolated from tail tip samples. Tails were peeled, minced into 1 cm pieces, placed on culture dishes, and incubated in MEF media for 5 days. Adult human fibroblasts were isolated from skin biopsy samples of healthy and Parkinson's disease patients^{22,23} provided from the "Cell Line and DNA Biobank from Patients affected by Genetic Diseases" (G. Gaslini Institute) and "Parkinson Institute Biobank" (Milan, <http://www.parkinson.it/dnabank.html>) of the Telethon Genetic Biobank Network (<http://www.biobanknetwork.org>). The informed consent as issued by the ICP Ethical committee was obtained by healthy and Parkinson's disease patients enrolled for the DNA and cell biobank collection.

Human skin samples were mechanically dissociated and plated on matrigel-coated dishes. Human fibroblasts were cultured as MEFs. Mouse and adult fibroblasts were grown in MEF media as well as human fetal lung fibroblasts IMR90 (ATCC). Mesencephalic dopaminergic primary cell cultures from TH-GFP mice were prepared as previously described²⁴. Mice were maintained at San Raffaele Scientific Institute Institutional mouse facility and experiments were performed in accordance with experimental protocols approved by local Institutional Animal Care and Use Committees.

Molecular cloning and viral infection. Complementary DNAs for the dopaminergic transcription factors were cloned into lentiviral vectors under the control of the tetracycline operator⁷. Replication-incompetent, VSVg-coated lentiviral particles were packaged in 293T cells. MEFs, IMR90 and adult mouse and human fibroblasts were infected in MEF media. 16–20 h after infection cells were switched into fresh MEF media containing doxycycline (2 mg ml⁻¹; Sigma). After 48 h medium was replaced with neuronal inducing media (DMEM/F12 (Invitrogen), 25 µg ml⁻¹ insulin (Sigma), 50 µg ml⁻¹ transferrin (Sigma), 30 nM sodium selenite, 20 nM progesterone (Sigma), 100 nM putrescine (Sigma) and penicillin/streptomycin (Sigma)) containing doxycycline. The medium was changed every 2–3 days for a further 10–22 days. For proliferation assay, MEFs were treated with a 48 h pulse of 10 µM BrdU.

Immunohistochemistry. For immunocytochemical analysis, 5 × 10⁴ mouse or human fibroblasts were plated on matrigel-coated glass coverslips the day before the infection. 10–28 days after viral infection cells were fixed for 20 min at room temperature (20–25 °C) in 4% paraformaldehyde in PBS, permeabilized for 30 min in PBS containing 0.1% Triton X-100 and 10% normal goat serum (NGS), and incubated overnight at 4 °C in PBS containing 10% NGS and primary antibodies. Then cells were washed three times with PBS and incubated for 2 h at room temperature with anti-rabbit or anti-mouse secondary antibodies Alexa Fluor-488 or Alexa Fluor-594 (1:500, Invitrogen). For immunohistochemical analysis P15 or adult mouse brains were fixed overnight at 4 °C with 4% paraformaldehyde, buffered in 30% sucrose and embedded in OCT. Frozen brains were sectioned into 15- or 40-µm-thick sections with a cryostat and processed for immunostaining. Sections were boiled for 3 min in 10 mM citrate buffer solution pH 6 for antigen retrieval and permeabilized for 1 h at room temperature in PBS containing 0.1% or 0.25% Triton X-100 and 10% NGS. Primary antibodies were as follows: mouse anti-TH (1:200, Millipore), rabbit anti-TH (1:200, Immunological Sciences), mouse anti-βIII-tubulin (1:500, Covance), rabbit anti-βIII-tubulin (TuJ1) (1:500, Covance), rabbit anti-VMAT2 (1:200, Chemicon), rat anti-DAT (1:500, Millipore), rabbit anti-D2 receptor (1:100, Millipore), rabbit anti-calbindin (1:200, Swant), rabbit anti-AADC (1:100, Novus Biologicals), rabbit anti-ALDH1A1 (1:200, Abcam), mouse anti-synaptotagmin I (1:200, Synaptic Systems), mouse anti-synapsin (1:200; Synaptic Systems), chicken anti-GFP (1:2,000, Molecular Probes), rat anti-BrdU (1:200, BD), mouse anti-MAP2 (1:500, Immunological Sciences), rabbit anti-Otx2 (1:100 R&D). β-Galactosidase staining was performed as previously described²⁵.

Statistical analysis. The total numbers of Th⁺ and TuJ1⁺ cells were quantified 12–24 days after infection. Cell counting was performed on ten fields from three replicates for each condition and normalized with the number of cells plated before the infection. Data were expressed as mean ± s.e.m.

RT-PCR. RNA was extracted from single cultures, using Trizol isolation system (Invitrogen) according to manufacturer's instructions. The yield and integrity of the RNA were determined by the spectrophotometric measurement of A260 and by agarose-gel electrophoresis, respectively. Total RNA was treated with DNase I (Qiagen) to prevent DNA contamination. Two micrograms of RNA were reverse

transcribed using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). One twentieth of the reverse transcribed cDNA was amplified in a 25 microlitres of reaction mixture containing Taq polymerase buffer (Fisher BioReagents), 0.2 mM dNTPs (Finnzymes OY), 0.4 micromolar of each primer, 1 U Taq polymerase (Fisher BioReagents). The primers used to amplify cDNA samples are listed in Supplementary Table 6.

Cell sorting, laser capture dissection and microarray analysis. TH-GFP-positive iDA cells were directly sorted in Trizol (Invitrogen) using the cell sorter FACSVantage s.e. DiVa (Becton Dickinson). Thus RNA was extracted as reported earlier and biotin-labelled cRNA was obtained using the Ovation kit (NuGEN). Labelled cRNA was hybridized (CBM geneexpression facility, SISSA) on Affymetrix Mouse Gene 1.0 ST Arrays, containing 35,557 probe sets corresponding to 28,853 genes. Hybridized arrays were stained and washed (GeneChip Fluidics Station 450) and scanned (GeneChip Scanner 3000 7G). Cell intensity values were computed using the Affymetrix GeneChip Operating Software (GCOS). Further data processing was performed in the R computing environment (<http://www.r-project.org/>) version 2.8.0 with BioConductor packages (<http://www.bioconductor.org/>). Robust Multi-Array Average (RMA) normalization was applied²⁶. Data were then filtered based on probe set intensity, so that only probe sets that had intensity values of > 50 in at least half the arrays were retained. Statistical analysis was performed with limma²⁷. *P* values were adjusted for multiple testing using Benjamini and Hochberg's method to control the false discovery rate²⁸. Genes with adjusted *P* values below 0.01 were considered to be differentially expressed. Furthermore, a fold-change threshold cutoff was set to focus on genes whose expression level changed at least 2 times. Data were analysed through DAVID Bioinformatics Resources v6.7 (refs 28, 29).

Gene expression profiles of adult A9 and A10 dopaminergic neurons were obtained as previously described³⁰. In brief, adult TH-GFP female mice were killed by cervical dislocation. The brains were rapidly cut to isolate the midbrain region, and immediately immersed in 1× zinc fixative (BD Pharmingen) for 4–6 h at +4 °C. Once fixed, tissues were moved to 30% sucrose in 1× zinc fixative solution at +4 °C overnight. Following inclusion in OCT, tissues were frozen in iso-pentan (Sigma) and percolated with liquid nitrogen. 14-mm cryosections were mounted on SuperFrost plus glass slides (Menzel-Gläser) and air dried. mDA A9 and A10 neurons (each one from three different mice) were isolated from cryosections by using a PALM LCM microdissection system (PALM Microlaser Technology). To facilitate detection of fluorescent neurons, a drop of 1× zinc fixative was applied to the section during cell selection. The sections were air dried, neurons were dissected and catapulted onto PALM adhesive caps (Zeiss). Total RNA from 2,500 pooled neurons was isolated by using the Nano RNA extraction kit (Stratagene) and contaminating genomic DNA was removed through on-column DNase digestion step. The common reference RNA was generated from three midbrain regions of age-matched female mice. Midbrain RNA was isolated using RNeasy Mini kit (Qiagen), followed by DNase treatment. RNA from dissected neurons and all midbrains was amplified and labelled by Ovation Pico kit, WT exon and Encore biotin labelling kit (Nugene), following manufacturer's instructions. Once prepared, each target was hybridized on MoExon 1.0ST GeneChip (Affymetrix). Statistical analysis was performed by oneChannelGUI R package. All hierarchical clusters were generated by TMEV software.

Bisulphite genomic sequencing. DNA from sorted TH-GFP⁺ reprogrammed MEFs was modified using the CpGenome modification kit (Chemicon) according to the manufacturer's recommendations. Thus *Th* and *Vmat2* promoters CpG-rich selected regions were amplified using the PCR primers listed in Supplementary Table 5.

HPLC. To quantify dopamine level in reprogrammed cells, cell pellets were homogenized in 100 µl 0.1 N HClO₄ and analysed by using HPLC with electrochemical detection (Alexis 100, Antec Leyden). To measure dopamine concentrations in the supernatants, cells were exposed to media with or without 50 mM KCl for 30 min, then 0.9 ml of supernatants were collected with the addition of 0.1 ml of 1 N HClO₄, filtered and analysed by HPLC. Dopamine was separated on a reverse-phase column (ALB-105, 3 µm, 50 × 1 mm) with a mobile phase consisting of 50 mM phosphate buffer, 8 mM KCl, 500 mg l⁻¹ octyl sodium sulphate, 0.1 mM EDTA, and 3% methanol (pH 6.0) at a flow rate of 50 µl min⁻¹. Dopamine was detected by a Decade II electrochemical detector equipped with micro VT-03 electrochemical flow cell and a 0.7-mm-diameter glassy carbon electrode (Alexis 100, Antec Leyden). The volume of the injection was 5 µl. The detection limit established as a 3:1 signal-to-noise ratio was below 0.5 nM.

Electrophysiology. Recordings were performed from reprogrammed mouse and human fibroblasts and primary mDA neurons. The mouse TH-GFP⁺ cells selected for the electrophysiological analysis were not so flat as the fibroblasts, and had several well-developed neurites. The human cells selected for the electrophysiological analysis also had neuron-like shapes with clearly distinguishable neurites by phase contrast microscopy. Only cells without signs of detachment

from the substrate were used for recordings. Cells were perfused continuously with HEPES-buffered saline (HBS) of the following composition (in mM): 140 NaCl, 5 KCl, 2 CaCl₂, 2 MgCl₂, 15 HEPES, and 25 glucose, pH 7.4. The patch pipette solution contained (in mM): 130 K-gluconate, 10 KCl, 0.5 CaCl₂, 15 HEPES, 5 EGTA, 8 NaCl, 2 MgATP, 0.3 Na₂GTP, and 10 glucose, pH adjusted to 7.2 with KOH. Action potentials were recorded in the on-cell and the current-clamp whole-cell configurations. A current was injected to have membrane potentials around -60 mV, and step currents from -50 pA to 40 pA were injected to elicit action potentials. Na⁺ currents and composite K⁺ currents were recorded in the voltage-clamp configuration by delivering voltage steps ranging from -100 mV to +20 mV in cells held at -60 mV. Delayed rectifier K⁺ currents were activated by 0.5 s voltage steps from -40 mV to +20 mV after a 0.5-s-long step to -40 mV. A-type K⁺ currents were isolated by subtraction of delayed rectifier K⁺ currents from those activated by voltage steps after a 0.5-s-long step to -100 mV. Recordings were performed using an EPC10 USB patch clamp amplifier and PATCHMASTER software (HEKA Elektronik). Data were digitized at 10 kHz and analysed with FITMASTER Software (HEKA Elektronik). Detection and measurements of action potentials were performed using MiniAnalysis software (Synaptosoft).

In vivo electrophysiology. Slices were obtained from transplanted mice at postnatal day 42. The brains were quickly removed from the skull in ice-cold artificial cerebrospinal fluid (ACSF) containing the following (in mM): 125 NaCl, 25 NaHCO₃, 2.5 KCl, 1.25 NaH₂PO₄, 2 CaCl₂, 1 MgCl₂, and 25 glucose, pH 7.4 (bubbled with 95% O₂ and 5% CO₂). Coronal slices (300-μm thick) were cut using a vibratome (VT1000S; Leica) and stored in ACSF at 25–28 °C. For recording, slices were transferred to a recording chamber continuously superfused with ACSF (1–2 ml min⁻¹ at 30–32 °C). Whole-cell recordings were performed in both current- and voltage-clamp configurations. Recording pipettes (3–5 MΩ of resistance) contained the following solution (in mM): 124 KH₂PO₄, 10 NaCl, 2 MgCl₂, 0.5 EGTA, 10 HEPES, 2 Na₂-ATP, 0.03 Na-GTP (pH 7.2, adjusted with KOH). Signals were sampled at 10 kHz, filtered at 2 kHz, and acquired using a MultiClamp 700A amplifier and pClamp 10 software (Molecular Devices).

Amperometric recording. Amperometry was used to detect the evoked catecholamine exocytosis from single cells^{15,31}. Carbon-fibre microelectrodes were fabricated from 5 μm carbon fibres (Goodfellow), inserted in a 1.2 × 0.68 mm glass capillary (A-M system, Sequim) and pulled with a PE-22 micropipette puller (Narishige). Electrodes were sealed by dipping in Epoxy resin (Epo-Tek 301, Epoxy Technology) and cured at 100 °C for 24 h. They were backfilled with 3 M KCl and trimmed to obtain a basal current between 140 and 180 nA. The electrodes' responses were tested by cyclic voltammetry and those with unstable cyclic voltammograms, when tested in a solution of 10 μM dopamine, were rejected. A voltage was applied to the carbon fibre using an EPC10 USB patch-clamp amplifier (HEKA Elektronik). The signal was low-pass filtered at 10 kHz using a 4-pole Bessel filter, digitalized at 50 kHz and digitally refiltered at 1–1,000 Hz. The latter resulted in slightly longer responses, but significantly improved the visualization of secretion events that was the only aim of these experiments. The electrode was positioned adjacent to individual cells and lowered to approach the somatodendritic domains of iDA cells³², using an Olympus BX51WI microscope with ×40 water immersion objective. To increase the signal-to-noise ratio, cells were pre-treated with 100 μM L-DOPA (Sigma-Aldrich) for 30 min although we were able to resolve single spike-like release events in two untreated cells. The experiments consisted of current recordings at +750 mV during a brief baseline period, during which cells were perfused with standard external medium containing 5 mM K⁺. It was then exchanged for a stimulation solution (25 mM K⁺), and amperometric signals were recorded for a further period of 7 min. Catecholamine secretion was apparent as discrete spike-like events, each corresponding to vesicular catecholamine release. Most events were detected during 2 min after 25 mM K⁺ stimulation, but occasional events were observed also during baseline recordings. No vesicular release of dopamine was recorded at the electrode placed adjacent to a cell when the applied potential was 0 mV or -750 mV, or at +750 mV when the electrode was placed remotely from cells.

FM4-64 assay. FM4-64 dye uptake experiments were performed as previously reported³³. Briefly, 21 days *in vitro* TH-GFP⁺ iDA cells were stimulated for 1 min with 55 mM KCl, in the presence of FM4-64 (10 μM). After FM4-64 loading, neuronal cells were washed and perfused for 10 min with warmed Krebs buffer (37 °C) supplemented with TTX (1 μM) and CNQX (10 μM). After live fluorescent FM4-64 signals were acquired, cells were fixed and immunostained for TH and SYT1.

Electron microscopy. For ultrastructural immunocytochemistry, 21 days *in vitro* infected MEFs were fixed in 2% glutaraldehyde in PBS, washed in PBS, postfixed in 2% OsO₄ in PBS, and embedded in Epon. Ultrathin sections prepared from these samples were analysed with electron microscope (H-7000; Hitachi).

Transplantation surgery. After 4 days of infection, TH-GFP MEFs were trypsinized and resuspended at 2 × 10⁵ cells μl⁻¹ in fresh prepared Krebs buffer containing the following (in mM): 126 NaCl, 2.5 KCl, 1.2 NaH₂PO₄, 1.2 MgCl₂, 2.1 CaCl₂, 11 glucose, 4.2 NaHCO₃, 1 HEPES, and 1% vital dye Fast Green. P1 mice pups were anaesthetized by hypothermia (4 min) and fixed to a support using band-aid. The skin and the skull overlying the lateral ventricle were opened over about 2 mm using an ophthalmic scalpel. Subsequently, the mouse was placed in a stereotaxic rig (Kopf) under a Hamilton syringe containing 2 μl of cell suspension. The syringe was placed over the incision, positioned at the level of the skull, then lowered into the lumen of the right lateral ventricle (LV; 2.5 mm) or in the somatosensory cortex (1.5 mm) and cell solution was injected. Mice were left on a 37 °C heating blanket for several minutes after surgical manipulation to avoid fatal hypothermia.

Stereological analysis. Three mice transplanted with reprogrammed cells were used for stereological analysis. Three weeks after transplantation mice were anaesthetized and killed by transcardiac perfusion with PBS followed by 4% paraformaldehyde. Brains were cryoprotected through incubation in an ice-cold solution of 30% sucrose in PBS and cut in coronal 40-μm-thick cryostat sections. From these sections, one systematic random series of sections was stained for GFP, so that sections were spaced at 7 section intervals (total of 16 sections per mouse). GFP immunoperoxidase staining was performed as described elsewhere³⁴. Cells were quantified using the assistance of the Stereo Investigator v 3.0 software (MicroBrightField) and a personal computer running the software connected to a colour video camera mounted on a Leica microscope^{35,36}. The motorized stage of the microscope allowed precise and well-defined movements along the x-, y- and z-axes. Images were first acquired with a CCD-IRIS colour video camera and the cerebral hemispheres were interactively delineated at low magnification on a video image of the section. Counting of cells was performed manually on every seventh section using a ×40 lens. To estimate the total number of GFP positive cells the total number of neurons counted on the sections was multiplied by seven.

22. Sironi, F. *et al.* α-Synuclein multiplication analysis in Italian familial Parkinson disease. *Parkinsonism Relat. Disord.* **16**, 228–231 (2010).
23. Sironi, F. *et al.* Parkin analysis in early onset Parkinson's disease. *Parkinsonism Relat. Disord.* **14**, 326–333 (2008).
24. Pruszek, J. *et al.* Isolation and culture of ventral mesencephalic precursor cells and dopaminergic neurons from rodent brain. *Curr. Prot. Stem Cell Biol.* Chapter 2, 2D.5.1–2D.5.21 (2009).
25. Broccoli, V. *et al.* The caudal limit of *Otx2* expression positions the isthmus organizer. *Nature* **401**, 164–168 (1999).
26. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
27. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article 3 (2004).
28. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).
29. Huang, D. W. *et al.* Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
30. Biagioli, M. *et al.* Unexpected expression of α- and β-globin in mesencephalic dopaminergic neurons and glial cells. *Proc. Natl Acad. USA* **106**, 15454–15459 (2009).
31. Pothos, E., Desmond, D. & Sulzer, D. L-3,4-Dihydroxyphenylalanine increases the quantal size of exocytotic dopamine release *in vitro*. *J. Neurochem.* **66**, 629–636 (1996).
32. Mundroff, M. L. & Wightman, R. M. Amperometry and cyclic voltammetry with carbon fiber microelectrodes at single cells. *Curr. Protoc. Neurosci.* Chapter 6, 6.14.1–6.14.22 (2002).
33. Menegon, A. *et al.* Protein kinase A-mediated synapsin I phosphorylation is a central modulator of Ca²⁺-dependent synaptic activity. *J. Neurosci.* **26**, 11670–11681 (2006).
34. Colasante, G. *et al.* Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. *J. Neurosci.* **28**, 10674–10686 (2008).
35. West, M. J. *et al.* Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.* **231**, 482–497 (1991).
36. Bacigaluppi, M. *et al.* Delayed post-ischaemic neuroprotection following systemic neural stem cell transplantation involves multiple mechanisms. *Brain* **132**, 2239–2251 (2009).

MicroRNA-mediated conversion of human fibroblasts to neurons

Andrew S. Yoo^{1†*}, Alfred X. Sun^{2*}, Li Li^{3,4,5*}, Aleksandr Shcheglovitov^{6*}, Thomas Portmann⁶, Yulong Li³, Chris Lee-Messer⁷, Ricardo E. Dolmetsch⁶, Richard W. Tsien³ & Gerald R. Crabtree¹

Neurogenic transcription factors and evolutionarily conserved signalling pathways have been found to be instrumental in the formation of neurons^{1,2}. However, the instructive role of microRNAs (miRNAs) in neurogenesis remains unexplored. We recently discovered that miR-9* and miR-124 instruct compositional changes of SWI/SNF-like BAF chromatin-remodelling complexes, a process important for neuronal differentiation and function^{3–6}. Nearing mitotic exit of neural progenitors, miR-9* and miR-124 repress the BAF53a subunit of the neural-progenitor (np)BAF chromatin-remodelling complex. After mitotic exit, BAF53a is replaced by BAF53b, and BAF45a by BAF45b and BAF45c, which are then incorporated into neuron-specific (n)BAF complexes essential for post-mitotic functions⁴. Because miR-9/9* and miR-124 also control multiple genes regulating neuronal differentiation and function^{5,7–13}, we proposed that these miRNAs might contribute to neuronal fates. Here we show that expression of miR-9/9* and miR-124 (miR-9/9*-124) in human fibroblasts induces their conversion into neurons, a process facilitated by *NEUROD2*. Further addition of neurogenic transcription factors *ASCL1* and *MYT1L* enhances the rate of conversion and the maturation of the converted neurons, whereas expression of these transcription factors alone without miR-9/9*-124 was ineffective. These studies indicate that the genetic circuitry involving miR-9/9*-124 can have an instructive role in neural fate determination.

During the course of exploring the roles of miR-9/9* and miR-124 (ref. 5), we noted that these miRNAs induced neuronal morphologies in cultured cells. To explore this effect in greater detail, we prepared a single lentiviral vector that expresses both precursors of miR-9/9* and miR-124 along with a turbo red fluorescent protein (tRFP) marker, and infected human neonatal foreskin fibroblasts (Supplementary Fig. 1). The fibroblast culture was free of neural progenitors, keratinocytes or melanocytes (Supplementary Figs 2–4). Remarkably, fibroblasts expressing miR-9/9*-124 showed a rapid reduction in proliferation, displayed neuron-like morphologies (Supplementary Fig. 5) and expressed MAP2, a marker of post-mitotic neurons, within 4 weeks after infection (Fig. 1a, left). This was owing to synergism between miR-9/9* and miR-124, as expressing these miRNAs separately did not lead to the appearance of MAP2-positive cells (Supplementary Fig. 6). In light of the low percentage of MAP2-positive cells obtained with miRNAs only (less than 5%, Fig. 1b), we began adding neurogenic transcription factors and found that *NEUROD2* (refs 14–17) was most effective at increasing the conversion frequency (Supplementary Fig. 7). We estimate that ~50% of these cells have acquired neuronal fates as indicated by MAP2 expression 30 days after infection (Fig. 1a, right, b). However, because cells detached, remained uninfected or died during the conversion process, a conservative estimate is that ~5% of the starting cells became neurons. Importantly, neither *NEUROD2* alone nor non-specific miRNA (miR-NS) could convert fibroblasts into neurons

(Fig. 1b), demonstrating the essential role of miR-9/9*-124 in this process. Synergism between miR-9/9* and miR-124 seemed to be crucial: expressing miR-9/9* and miR-124 individually with *NEUROD2* failed to produce MAP2-positive cells (Supplementary Fig. 6). Using EdU-incorporation, we found that miR-9/9*-124-infected fibroblasts had exited the cell cycle 1 week after infection (Supplementary Fig. 8), consistent with the anti-proliferative role of these miRNAs⁵. Lastly, immunostains indicated that the induced neurons expressed *SCN1a*, a key contributor to neuronal excitability, as well as synapsin 1 and NMDA receptor 1 (Fig. 1c).

Using whole-cell patch recording, we found that injecting depolarizing current in induced neurons (cultured up to 8 weeks) could consistently trigger single action potentials and in some cases, repetitive firing (Fig. 1d). Moreover, their resting membrane potential (-34.1 ± 1.7 mV; Supplementary Fig. 9) was significantly more negative than that of control fibroblasts (-20.4 ± 0.6 mV, $n = 4$). Applying a series of voltage steps to the induced cells evoked large inward currents closely followed by outward currents, which were not observed in the fibroblasts (Supplementary Fig. 10). Importantly, adding 1 μ M tetrodotoxin (TTX) completely and reversibly blocked the initial inward current, confirming that the current was due to voltage-gated sodium channels (Fig. 1e), as would be expected from the current-voltage ($I-V$) curve of inward currents (Fig. 1f, left). The $I-V$ curve of outward currents showed the characteristics of voltage-gated potassium channels in neurons (Fig. 1f, right). Moreover, some of these cells exhibited postsynaptic currents, which could be reversibly blocked by 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[*f*]quinoxaline-2,3-dione (NBQX) and 2-amino-5-phosphonopentanoic acid (APV) (Supplementary Fig. 11).

We examined the ability of cells converted by miR-9/9*-124-*NEUROD2* to elicit a stimulation-dependent calcium influx using the calcium indicator Fluo2. Field stimulation triggered calcium influx that could be abolished by adding TTX (Fig. 1g) or 200 μ M Cd^{2+} (Supplementary Fig. 12), demonstrating the ability of converted cells to support activity-dependent Ca^{2+} influx through voltage-gated Ca^{2+} channels without any requirement for a pre-pulse. Activity-dependent uptake and release of the lipophilic dye FM1-43 was used to evaluate the ability to form functional presynaptic terminals¹⁸. We found that the induced cells were able to take up and release FM dyes in a stimulation-dependent and Ca^{2+} -dependent manner (Fig. 1h).

Because the miR-9/9*-124-*NEUROD2*-induced cells only occasionally showed repetitive action potentials, we sought to optimize the maturation of the cells by introducing additional neurogenic factors. Because *ASCL1* and *MYT1L* were previously shown to be important for converting mouse embryonic fibroblasts into functionally mature neurons¹⁹, we expressed miR-9/9*-124 together with *NEUROD2*, *ASCL1* and *MYT1L* (DAM). We found that the miR-9/9*-124-DAM-converted cells were positive for MAP2 expression in approximately

¹Howard Hughes Medical Institute and the Departments of Developmental Biology and of Pathology, Stanford University, Stanford, California 94305, USA. ²Program in Cancer Biology, Stanford University, Stanford, California 94305, USA. ³Department of Molecular and Cellular Physiology, Stanford University, Stanford, California 94305, USA. ⁴Medical Scientist Training Program, Stanford University, Stanford, California 94305, USA. ⁵Neuroscience Program, Stanford University, Stanford, California 94305, USA. ⁶Department of Neurobiology, Stanford University, Stanford, California 94305, USA.

⁷Department of Neurology, Stanford University, Stanford, California 94305, USA. [†]Present address: Department of Developmental Biology, Washington University in St Louis, St Louis, Missouri 63110, USA.

*These authors contributed equally to this work.

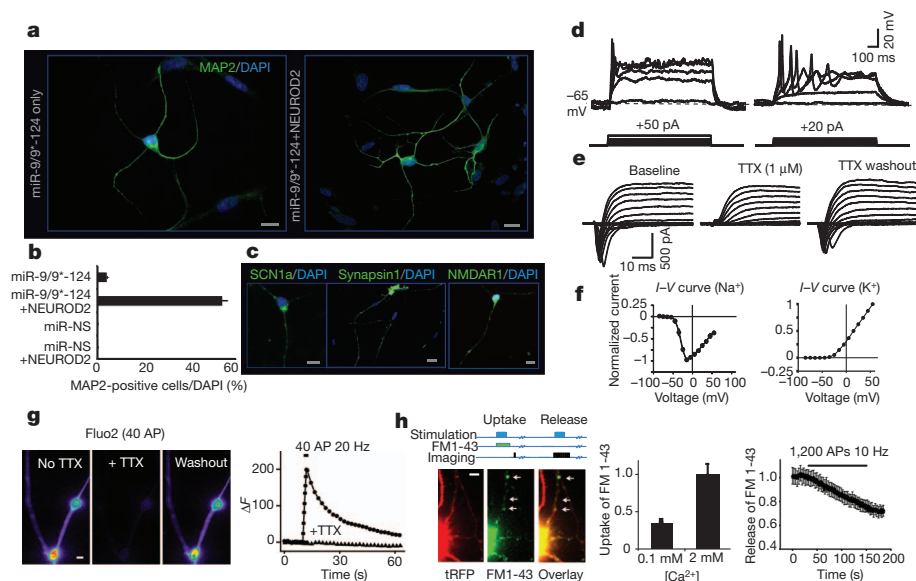


Figure 1 | miRNA-induced transformation of human fibroblasts. **a**, MAP2 expression in miR-9/9*-124 only (left) and miR-9/9*-124-NEUROD2 (right) converted cells. DAPI, 4',6-diamidino-2-phenylindole. Scale bar, 20 μ m. **b**, Quantification of MAP2-positive cells with processes at least three times the length of the cell body from ten random fields. The graph represents the percentage of MAP2-positive cells over DAPI-positive cells. miR-9/9*-124 only: $n = 558$; miR-9/9*-124-NEUROD2: $n = 658$ cells. The error bars are s.e.m. MAP2 signal was undetectable in fibroblasts infected with miR-NS, with or without NEUROD2. Scale bar, 20 μ m. **c**, Expression of SCN1a, synapsin 1 and NMDAR1 in miR-9/9*-124-NEUROD2-converted cells. Scale bar, 20 μ m. **d**, Representative traces of action potentials recorded in current clamp in miR-9/9*-124-NEUROD2-converted cells. Twelve out of 22 cells showed single action potentials and 2 cells showed repetitive firing. **e**, A representative example of a series of voltage steps applied to an miR-9/9*-124-NEUROD2-induced neuron held at -70 mV. An inward current was observed, blocked by 1μ M TTX, and reversed after TTX washout ($n = 6$). **f**, I - V curve for the peak

inward (left) and outward (right) currents. **g**, An example of Ca^{2+} influx in induced neurons as measured by Fluo2-AM imaging. Images show the peak Fluo2-AM signal on stimulation before and during TTX application, and after TTX washout, respectively. The graph represents the changes in Fluo2 signal over time (circles, no TTX; triangles, with TTX). AP, action potential. Field stimulation is indicated by the black bar. Scale bar, 2 μ m. **h**, An example of vesicle recycling measured by FM1-43 imaging in induced neurons. Top diagram illustrates the protocol of FM uptake or release experiments. The images represent the typical FM1-43 dye uptake signal (middle) in a converted cell marked by tRFP (left). The left graph shows the measurement of FM1-43 loading in $2 \text{ mM } \text{Ca}^{2+}$, which was significantly reduced in low Ca^{2+} concentration (0.1 mM). The right graph quantifies FM1-43 release (de-staining) during stimulations. FM1-43 signal was measured from $n > 600$ boutons (arrows indicate examples of boutons) from 4 cultures. All error bars are s.e.m. In some cases, the s.e.m. is too small to be resolved. Scale bar, 4 μ m.

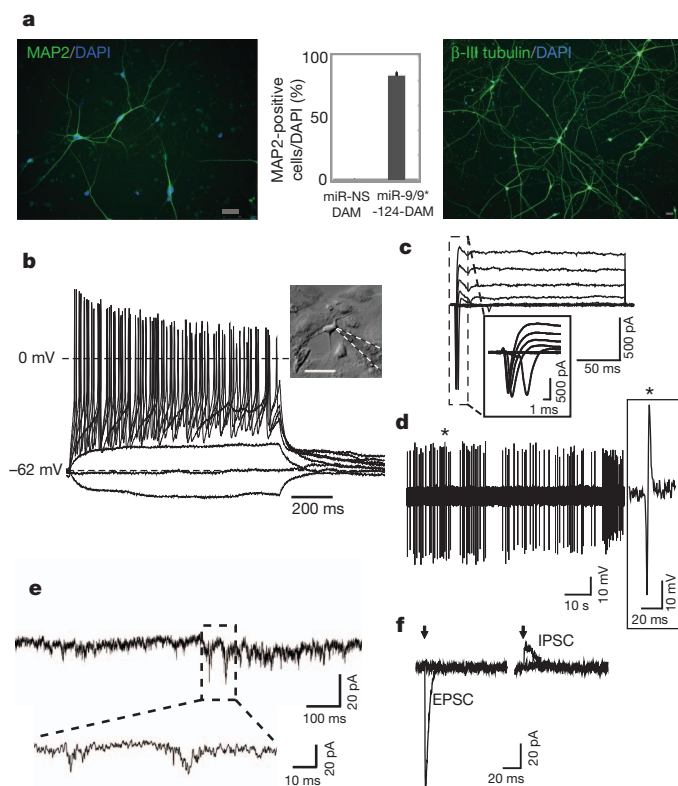


Figure 2 | Additional neural factors enhance the conversion to neurons. **a**, MAP2 (left) and β -III tubulin (right) immunostaining of miR-9/9*-124-DAM-converted cells. Scale bar, 40 μ m. The graph represents the percentage of MAP2-positive cells over DAPI-positive cells. The error bars are s.e.m.; $n = 150$ cells. **b**, A representative current clamp recording from a cell with typical neuronal morphology (see inset; scale bar, 50 μ m). Voltage deflections were elicited by somatic current injections of various amplitudes ($\Delta = 5$ pA). **c**, A representative voltage clamp recording of the net current at various membrane potentials (-40 to $+20$ mV, $\Delta V = 10$ mV, $V_{\text{hold}} = -90$ mV). **d**, A representative trace of spontaneously active cells recorded in cell-attached mode. **e**, A representative trace demonstrating spontaneous EPSCs. **f**, Representative traces of evoked postsynaptic currents (left, EPSC; right, IPSC) obtained in response to local field stimulation with single current pulses (1 ms) of various amplitudes (left, 0.25 and 0.3 mA; right, 0.3 and 0.4 mA) at different membrane holding potentials (left, -70 mV; right, 0 mV). The arrows indicate the time when stimulation was applied. Stimulation artefacts were eliminated for clarity.

80% of the cells remaining on the coverslips (Fig. 2a, representing ~10% of the initially plated cells), and showed extensive neurite outgrowth as illustrated by β -III tubulin staining (Fig. 2a, right). In addition, miR-9/9*-124-DAM resulted in complete exit from cell cycle as assayed by EdU pulsing for 4 days (0/176 positive) whereas nearly all control cells were positive for EdU (97/107). Importantly, DAM factors with miRNA-NS failed to produce neurons as assayed by MAP2 staining (Supplementary Fig. 13). About 80% of the miR-9/9*-124-DAM-converted cells were able to fire repetitive action potentials in response to current injections, and showed typical sodium and potassium currents during voltage clamp depolarizations (Fig. 2b–c and Supplementary Figs 9 and 14). Among recorded cells, we also observed spontaneously active cells (2/21) (Fig. 2d). Spontaneous excitatory postsynaptic currents (EPSCs) were seen in 10 out of 14 induced cells (Fig. 2e) without co-cultured primary neurons (Supplementary Figs 9 and 15a). Furthermore, the induced cells exhibited evoked EPSCs and inhibitory postsynaptic currents (IPSCs) in response to local stimulation (Fig. 2f). Importantly, neuronal identity was stable after the removal of exogenous expression of miR-9/9*-124 and DAM after 3 weeks of induction, as they still stained positive for SV2 and synapsin 1 (Supplementary Fig. 16).

We next performed single-cell analysis to characterize the types of neurons in miR-9/9*-124-DAM-induced cells. From randomly collected single cells 4 weeks post-infection, we analysed a total of 45 induced neurons (based on *MAPT* and *TUBB3* co-expression) for genes expressed in different types of neurons. We found that most induced cells were positive for genes expressed in cortical layers (Fig. 3a and Supplementary Fig. 17). Interestingly, we did not detect a peripheral nervous system marker (peripherin) or dopaminergic/noradrenergic markers (*DDC*, *TH*). Striatal markers (*DLX5* and *DARPP32* (also known as *PPP1R1B*)), the serotonergic marker *5HT-2C* (also known as *HTR2C*), and cerebellar genes (*PCP2*, *GRP*, *TPM2*)

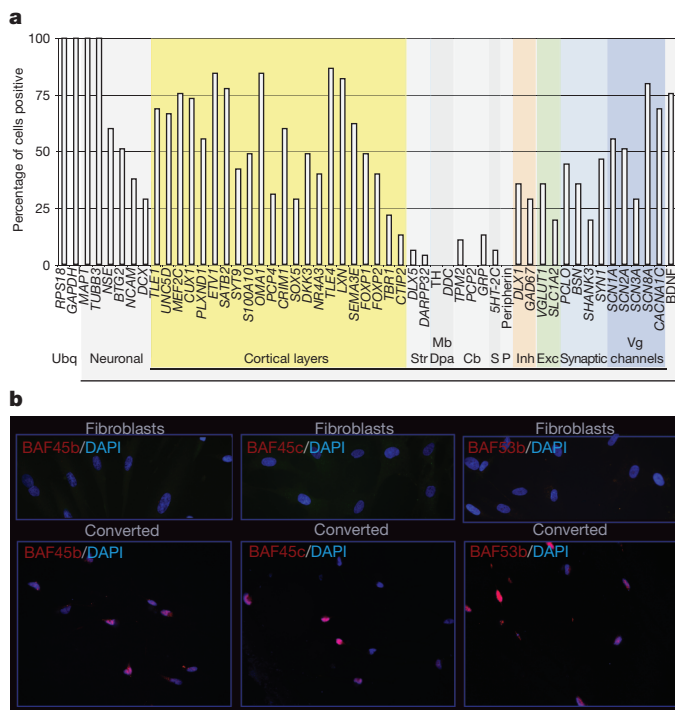


Figure 3 | Characterization of induced neurons and nBAF subunit expression. **a**, Multiplex quantitative polymerase chain reaction (qPCR) of 45 induced neurons for genes of specific brain regions and cell types. Cb, cerebellum; Dpa, dopaminergic; Exc, excitatory; Inh, inhibitory; Mb, midbrain; P, peripheral nervous system; S, serotonergic; Str, striatum; Ubq, ubiquitous; Vg, voltage gated. **b**, Fibroblasts stained negative for neuron-specific subunits of BAF complexes (top) whereas induced neurons expressed BAF45b, BAF45c and BAF53b (bottom). Scale bar, 20 μ m.

were expressed in only a small number of cells. The miR-9/9*-124-DAM-induced cells seemed to be heterogeneously excitatory (*VGLUT1* (also known as *SLC17A7*) and *SLC1A2*) and inhibitory (*GAD67* (also known as *GAD1*) and *DLX1*) (Fig. 3a), which is further supported by immunostains of VGLUT1 and GABA (Supplementary Fig. 18). Moreover, by 4 weeks post-infection, the induced cells already expressed genes important for synaptic structure and function, including *SYN1*, *BSN*, *PCLO* and *SHANK3* (Fig. 3a).

miR-9* and miR-124 target separate sites in the 3' untranslated region (UTR) of BAF53a, a subunit of BAF complexes resulting in repression of BAF53a and activation of BAF53b, which is involved in an evolutionarily conserved program of neural development²⁰. Remarkably, we found that all of the nBAF subunits (BAF53b, 45b and 45c) were induced in the converted cells (Fig. 3b). In embryonic stem cells, BAF complexes function across the genome at several thousand sites to control placement of polycomb repressive complex 2 and the H3K27me3 repressive mark^{21,22}. Hence, one role of the miRNAs might be to induce stable epigenetic changes involving polycomb function across the genome.

In addition to BAF53a, miR-9/9* and miR-124 also target other genes essential for neurogenesis and neuronal functions²³ including components of the REST complex such as REST and CoREST^{8,9,24–26}, and PTBP-1 (ref. 7). We found that human fibroblasts expressed BAF53a, which could be repressed by miR-9/9*-124 (Supplementary Fig. 19). However, prolonging the expression of BAF53a only incompletely blocked neuronal conversion of fibroblasts, as assayed by MAP2 staining (data not shown). Prolonging the expression of REST, CoREST or PTBP1 yielded similar results (data not shown). These findings indicate that in inducing cell fate transformations, the miRNAs miR-9/9* and miR-124 operate programmatically on multiple targets.

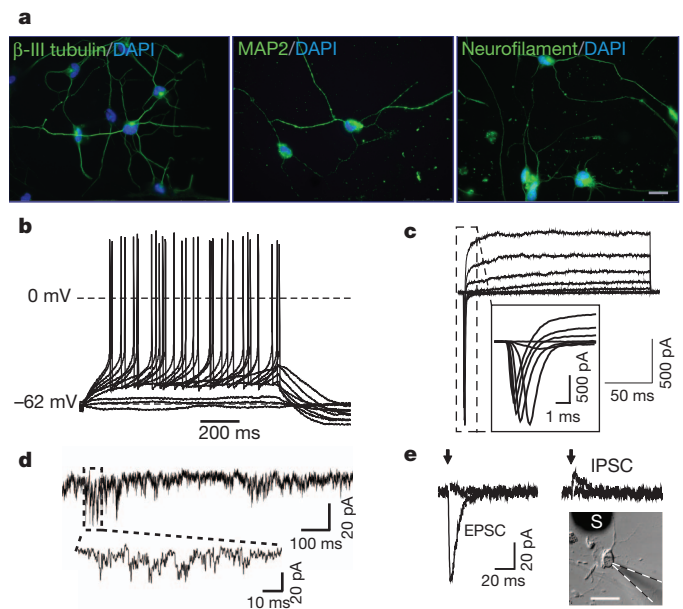


Figure 4 | Conversion of adult fibroblasts by miR-9/9*-124-DAM. **a**, Immunostaining of β -III tubulin (left), MAP2 (middle) and neurofilament (right) in dermal fibroblasts of a 30-year-old individual converted by miR-9/9*-124-DAM. Scale bar, 20 μ m. **b**, A representative current clamp recording. Voltage deflections were elicited by somatic current injections of various amplitudes ($\Delta = 2$ pA). **c**, A representative voltage clamp recording of the net current at various membrane potentials (-40 to $+20$ mV, $\Delta V = 10$ mV, $V_{\text{hold}} = -90$ mV). **d**, A representative trace of spontaneous EPSCs. **e**, Representative traces of evoked postsynaptic currents obtained from converted adult cells (see inset; scale bar, 50 μ m) in response to local field stimulation with single current pulses (1 ms) of various amplitudes (left, 0.2 and 0.25 mA; right, 0.4 and 0.5 mA). Evoked EPSPs and IPSCs were recorded at -70 mV and $+30$ mV in the presence of picrotoxin and NBQX/APV, respectively. S, stimulator.

Lastly, we asked whether our approach could be effective in converting adult fibroblasts. We found that adult human dermal fibroblasts (from a 30-year-old female) could be converted into neurons (Fig. 4a), albeit more slowly. Recordings from adult cells converted by miR-9/9*-124-DAM 6 weeks after infection showed that they were able to generate action potentials (Fig. 4b). They also demonstrated typical voltage-gated sodium and potassium currents (Fig. 4c), spontaneous EPSCs (Fig. 4d and Supplementary Fig. 15b) and evoked EPSCs and IPSCs (Fig. 4e) without co-cultured primary neurons.

Our studies show that activating a neural developmental regulatory circuit involving miRNAs in human fibroblasts can surprisingly induce their conversion into neurons, indicating an instructive role for this circuitry in human neurogenesis. In our study, neurogenic transcription factors, delivered either singly (*NEUROD2*), or in combination (*NEUROD2*, *ASCL1* and *MTYL1*), seem to function synergistically with the neurogenic activities of miR-9/9*-124. This raises the possibility of inducing various types of neurons using miR-9/9*-124 together with different sets of transcriptional factors.

METHODS SUMMARY

Transduction of human fibroblasts. A synthetic cluster of miR-9/9* and miR-124 validated previously to express miR-9* and miR-124 (ref. 5) was inserted downstream of tRFP in the pLentiR lentiviral construct carrying a puromycin selection cassette (Open Biosystems) driven by either a CMV promoter or a doxycycline-responsive promoter. A non-silencing sequence, which produces miRNA-NS, was used as a control (Open Biosystems). Each transcription factor was cloned downstream of the EF1 α promoter in a separate lentiviral construct. Typically, infected human fibroblasts were maintained in fibroblast media for 3–4 days before selection with appropriate antibiotics in Neuronal Media (ScienCell) supplemented with VPA (1 mM) and basic FGF (20 ng ml⁻¹). dbcAMP (500 μ M) was added 10 days later to enhance cell survival. Human BDNF and NT3 (10 ng ml⁻¹; Peprotech) were added to the media after 3–4 weeks. Media were changed every 4 days.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 July 2010; accepted 27 June 2011.

Published online 13 July 2011.

- Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Hansen, D. V., Rubenstein, J. L. & Kriegstein, A. R. Deriving excitatory neurons of the neocortex from pluripotent stem cells. *Neuron* **70**, 645–660 (2011).
- Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
- Wu, J. et al. Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron* **56**, 94–108 (2007).
- Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* **460**, 642–646 (2009).
- Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the words of chromatin regulation. *Cell* **136**, 200–206 (2009).
- Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
- Packer, A. N., Xing, Y., Harper, S. Q., Jones, L. & Davidson, B. L. The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. *J. Neurosci.* **28**, 14341–14346 (2008).
- Visvanathan, J., Lee, S., Lee, B., Lee, J. W. & Lee, S. K. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes Dev.* **21**, 744–749 (2007).
- Cheng, L. C., Pastrana, E., Tavazoie, M. & Doetsch, F. miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nature Neurosci.* **12**, 399–408 (2009).

- Krichevsky, A. M., Sonntag, K. C., Isacson, O. & Kosik, K. S. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem Cells* **24**, 857–864 (2006).
- Maierano, N. A. & Mallamaci, A. Promotion of embryonic cortico-cerebral neurogenesis by miR-124. *Neural Develop.* **4**, 40 (2009).
- Tang, X. et al. A simple array platform for microRNA analysis and its application in mouse tissues. *RNA* **13**, 1803–1822 (2007).
- Lin, C. H. et al. The dosage of the neuroD2 transcription factor regulates amygdala development and emotional learning. *Proc. Natl Acad. Sci. USA* **102**, 14877–14882 (2005).
- McCormick, M. B. et al. *neuroD2* and *neuroD3*: distinct expression patterns and transcriptional activation potentials within the *neuroD* gene family. *Mol. Cell. Biol.* **16**, 5792–5800 (1996).
- Olson, J. M. et al. *NeuroD2* is necessary for development and survival of central nervous system neurons. *Dev. Biol.* **234**, 174–187 (2001).
- Ince-Dunn, G. et al. Regulation of thalamocortical patterning and synaptic maturation by *NeuroD2*. *Neuron* **49**, 683–695 (2006).
- Ryan, T. A. et al. The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
- Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Parrish, J. Z., Kim, M. D., Jan, L. Y. & Jan, Y. N. Genome-wide analyses identify transcription factors required for proper morphogenesis of *Drosophila* sensory neuron dendrites. *Genes Dev.* **20**, 820–835 (2006).
- Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl Acad. Sci. USA* **106**, 5187–5191 (2009).
- Ho, L., Miller, E. L., Ronan, J. L., Ho, W. Q., Jothi, R. & Crabtree, G. R. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signaling and by regulating polycomb function. *Nature Cell Biol.* (in the press).
- Coolen, M. & Bally-Cuif, L. MicroRNAs in brain development and physiology. *Curr. Opin. Neurobiol.* **19**, 461–470 (2009).
- Wu, J. & Xie, X. Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome Biol.* **7**, R85 (2006).
- Laneve, P. et al. A minicircuitry involving REST and CREB controls miR-9-2 expression during human neuronal differentiation. *Nucleic Acids Res.* **38**, 6895–6905 (2010).
- Andres, M. E. et al. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl Acad. Sci. USA* **96**, 9873–9878 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Graef and A. Cho for helpful suggestions and reagents, A. Kuo and W. Ho for technical help, and X. Bao and P. Khavari for their generous gift of reagents. A.S.Y. is a fellow of the Helen Hay Whitney Foundation. A.X.S. is funded by the Agency of Science, Technology and Research of Singapore (A*STAR). L.L. is supported by the Stanford Medical Scientist Training Program, National Institutes of Mental Health (NIMH) F30MH093125, and the Frances B. Nelson predoctoral fellowship. A.S. is supported by the CIRM post-doctoral fellowship. T.P. is supported by a Swiss National Science Foundation SNSF fellowship for advanced researchers (PA00P3_134196). R.E.D. is supported by the NIH Director's Award, and awards from the Simon's Foundation and the CIRM. R.E.D. is also grateful for funding from B. and F. Horowitz, M. McCaffrey, B. and J. Packard, P. Kwan and K. Wang. R.W.T. is supported by grants from the Simons, Mathers and Burnett Family Foundations. This work was supported by grants from the Howard Hughes Medical Institute (G.R.C.) and the NIH (HD55391, AI060037 and NS046789 to G.R.C., and NS24067, GM58234 and MH064070 to R.W.T.).

Author Contributions A.S.Y., A.X.S., and G.R.C. generated the hypotheses and designed experiments. A.S.Y. and A.X.S. performed experiments, generated data in all figures and Supplementary Data. A.S. and L.L. designed and performed experiments for Figs 1, 2 and 4 and Supplementary Data. T.P. designed and performed experiments in Fig. 3a. Y.L. generated data presented in Fig. 1. C.L.-M. performed experiments for Supplementary Data. A.S.Y., A.X.S., L.L., A.S., Y.L., T.P., R.W.T., R.E.D. and G.R.C. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.R.C. (crabtree@stanford.edu) or A.S.Y. (yooa@wustl.edu).

METHODS

Plasmid construction and viral preparation. We have previously constructed a synthetic cluster that expresses the precursors of both miR-9/9* (NCBI and miRBASE accession numbers MIMAT0000441 and MIMAT0000442) and miR-124 (accession number MIMAT0000422) and validated its ability to generate mature miRNAs of both. Here we cloned this cluster downstream of a rFP marker into pLemiR (Open Biosystems), driven by either a CMV promoter or a doxycycline-responsive promoter. A non-silencing sequence in pLemiR (miR-NS, which produces non-specific miRNA) was used as a control (Open Biosystems). cDNA of each neural transcription factor used in this study was cloned downstream of the EF1 α promoter in a separate lentiviral construct with blasticidin or neomycin selection. For doxycycline experiments, human fibroblasts were first infected with a lentiviral construct expressing rtTA under the EF1 α promoter and stably selected with hygromycin. Infectious lentiviruses were collected 36–60 h after transfection of Lenti-X 293T cells (Clontech) with appropriate amounts of lentiviral vectors, psPAX2 and pMD2.G (Addgene) using Fugene HD (Roche).

Cell culture. All fibroblast cultures (human neonatal foreskin fibroblasts (ATCC, PCS-201-010) and adult dermal fibroblast (ScienCell)) were maintained in fibroblast media (Dulbecco's Modified Eagle Medium; Invitrogen) containing 10% fetal bovine serum (FBS; Omega Scientific), β -mercaptoethanol (Sigma-Aldrich), non-essential amino acids, sodium pyruvate, GlutaMAX, and penicillin/streptomycin (all from Invitrogen). The day before lentiviral infection, human fibroblasts were seeded onto gelatin-coated 24-well tissue culture dishes (MidSci). Next day, cells were infected with filtered viral supernatants in the presence of polybrene ($8 \mu\text{g ml}^{-1}$) overnight. Fresh media were then replaced for 2–3 days with appropriate antibiotics to select for infected cells. Four days after infection, the media was changed to Neuronal Media (ScienCell) supplemented with VPA (1 mM), basic FGF (20 ng ml^{-1}) with media changes every 4 days. Our experiments to optimize the conversion protocol indicated that VPA and bFGF are beneficial for conversion for the first 2–3 weeks. We also added dbcAMP (0.5 mM, Sigma) after 2 weeks to the media as we found it enhanced cell survival. Human BDNF and NT3 (10 ng ml^{-1} , Peprotech) were added to the media after 3–4 weeks to promote the survival of the induced cells. To facilitate immunostaining and electrophysiological studies in some experiments, cells were trypsinized (0.05% trypsin, Invitrogen) at about 10 days after infection and re-plated onto poly-ornithine (Sigma-Aldrich)/laminin (Roche)/fibronectin (Sigma-Aldrich)-coated glass coverslips.

Immunofluorescence. The following antibodies were used for the immunofluorescence studies: mouse anti-MAP2 (Sigma-Aldrich, 1:750), chicken anti-MAP2 (Abcam, 1:30,000), mouse anti- β -III tubulin (Covance, 1:30,000), rabbit anti-VGLUT1 (Synaptic Systems, 1:2,000), rabbit anti-SCN1a (Abcam, 1:1,000), rabbit anti-NMDAR1 (Abcam, 1:2,000), rabbit anti-neurofilament 200 (Sigma-Aldrich, 1:2,000), mouse anti-SV2 (Developmental Studies Hybridoma Bank, 1:100), rabbit anti-GABA (Sigma, 1:2,000) and rabbit anti-synapsin1 (Cell Signaling, 1:200). Antibodies against BAF subunits were generated in our laboratory and used as the following concentrations: BAF45b (1:250), BAF45c (1:1,000) and BAF53b (1:500). The secondary antibodies were goat anti-rabbit or mouse IgG conjugated with Alexa-488, -568 or -647 (Invitrogen). For SCN1a and BAF53b staining, biotinylated secondary antibodies were detected using TSA amplification kit (Invitrogen). EdU incorporation assay was performed according to the manufacturer's protocols (Invitrogen). Images were captured using a Leica DM5000B microscope with Leica Application Suite (LAS) Advanced Fluorescence 1.8.0 and a Leica DMI4000B microscope with LAS v.2.8.1.

Electrophysiology. Recordings were performed on fibroblasts 5–8 weeks after infection for both miR-9/9*-124-NEUROD2 and miR-9/9*-124-DAM converted cells, which were co-cultured with mouse glia. Data were acquired in whole-cell mode at room temperature (25°C) using an Axopatch 200B amplifier (Molecular Devices) or EPC 10 amplifier (HEKA) and sampled at 5 kHz with a 2 kHz low-pass filter. Recording pipette resistance was 2–6 M Ω . Intrinsic neuronal properties were studied using the following solutions (in mM): extracellular, 140 NaCl, 2.5 KCl, 2.5 CaCl₂, 2 MgCl₂, 1 NaH₂PO₄, 20 glucose, 10 HEPES, pH 7.4; intracellular, 120 KGluc, 20 KCl, 4 NaCl, 4 Mg₂ATP, 0.3 NaGTP, 10 Na₂PCr, 0.5 EGTA, 10 HEPES, pH 7.25. Synaptic activity was measured using the same extracellular solution, supplemented with 50 μM APV, and the following intracellular solution (in mM):

135 CsMeS, 5 CsCl, 10 HEPES, 0.5 EGTA, 1 MgCl₂, 4 Mg₂ATP, 0.4 NaGTP, 5 QX-314, pH 7.4 CsOH. Correspondingly, EPSCs were measured at -70 mV whereas IPSCs were measured at either 0 or $+30 \text{ mV}$ (with 10 μM NBQX and 50 μM APV, E_{Cl^-} approximately -79 mV). Evoked postsynaptic currents were elicited by a stimulating electrode (CBAEC75, FHC) positioned 100–150 μm aside from the cell soma through which brief (1 ms) unipolar current pulses of various amplitudes (0.1–0.9 mA, $\Delta = 0.05$ –0.1 mA) were applied. Recordings were filtered at 2 kHz and digitally sampled at 10 kHz. Data were collected and initially analysed with Clampfit 10 or the Patchmaster software (HEKA). Further analysis was performed using IgorPRO and MS Excel. Series resistance was left uncompensated owing to the fragility of the cells, but was corrected in the current clamp calculations. The liquid junction potential was calculated to be 15 mV (Clampfit) and corrected in calculating resting membrane potentials according to published methods²⁷.

Calcium imaging. Cells were loaded with Fluo2-AM (5 μM , TEFLABS) in Tyrode solution (150 mM NaCl, 4 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 10 mM glucose, 10 mM HEPES, 310–315 mOsm, with pH at 7.35) for 30 min in a 37°C incubator. After two washes with Tyrode, cells were imaged using a filter cube (excitation $470 \pm 20 \text{ nm}$ and emission $535 \pm 50 \text{ nm}$). In some cases, 1 μM TTX or 200 μM CdCl₂ was superfused. All images were converted to TIFF files and analysed off-line with Metamorph or ImageJ. All error bars represent s.e.m. For analysis of FM-positive puncta, 1.3 μm in diameter regions of interest were used to cover functional boutons. Photobleaching was corrected by fitting the pre-stimulation baseline to a linear curve.

FM1-43 imaging. Cells were superfused with Tyrode solution. Switching of superfusion solution was carried out with a precision of $<2 \text{ s}$. Solutions contained 10 μM NBQX and 50 μM D-APV (Tocris Bioscience) to prevent possible recurrent activity and synaptic plasticity. All experiments were performed at room temperature and neurons were stimulated with platinum electrodes. Putative presynaptic boutons were stained with 8 μM FM1-43 (Molecular Probes) using field stimulation for 120 s at 10 Hz, followed by 60 s without stimulation to maximize the loading. In some experiments, 0.1 mM CaCl₂ was used to test the calcium dependency. After 10 min of washing with dye-free Tyrode's solution, individual boutons were de-stained by field stimulation. FM1-43 dyes were excited at 470 nm (D470-40x; Chroma) and their emission was collected at 535 nm (535/50m). tRFP was excited at 535 nm (535/50ex) and its emission was collected at 580 nm (580 lp). All images were taken at a frame rate of 1–3 Hz by a Cascade 512B camera.

Single-cell qPCR. Single cells were collected by clone FACS sorting using a BD influx sorter (BD Biosciences) in 10 μl of a pre-amplification mix containing 40 nM of all primers for genes of interest, and the following components of the CellsDirect One-Step qRT-PCR Kit (Invitrogen): 2 \times Reaction Mix, SuperScript III RT/Platinum Taq Mix. After sorting, samples were reverse transcribed and pre-amplified for 18 cycles. Pre-amplified samples were diluted (3 \times) with TE buffer and stored at -20°C . Sample and assay (primer pairs) preparation for 96.96 Fluidigm Dynamic arrays was done according to the manufacturer's recommendation. Briefly, sample was mixed with 20 \times DNA binding dye sample loading reagent (Fluidigm), 20 \times EvaGreen (Biotium) and TaqMan Gene Expression Master Mix (Applied Biosystems). Assays were mixed with 2 \times assay loading reagent (Fluidigm) and TE to a final concentration of 5 μM . The 96.96 Fluidigm Dynamic Arrays (Fluidigm) were primed and loaded on an IFC Controller HX (Fluidigm) and qPCR experiments were run on a Biomark System for Genetic Analysis (Fluidigm). Data were collected and analysed using the Fluidigm Real-Time PCR Analysis software (v.2.1.3 and v.3.0.2). Melting curves were used to determine specificity of each reaction. Further data analysis was performed using Microsoft Excel. In addition to collected single-cell material, every experiment contained samples for four standard dilutions of a mixed human cDNA library. The collected cells were confirmed based on *RSG18* (18S small ribosomal subunit) and *GAPDH* co-expression. Of these, induced neurons were identified by co-expression of two general neuronal genes *MAPT* and *TUBB3* for further analysis of genes specific to brain regions and cell types.

27. Barry, P. H. JPCalc, a software package for calculating liquid junction potential corrections in patch-clamp, intracellular, epithelial and bilayer measurements and for correcting junction potential measurements. *J. Neurosci. Methods* **51**, 107–116 (1994).

Crystal structure of the human centromeric nucleosome containing CENP-A

Hiroaki Tachiwana^{1*}, Wataru Kagawa^{1*}, Tatsuya Shiga^{1*}, Akihisa Osakabe¹, Yuta Miya¹, Kengo Saito¹, Yoko Hayashi-Takanaka², Takashi Oda³, Mamoru Sato³, Sam-Yong Park⁴, Hiroshi Kimura² & Hitoshi Kurumizaka¹

In eukaryotes, accurate chromosome segregation during mitosis and meiosis is coordinated by kinetochores, which are unique chromosomal sites for microtubule attachment^{1,2}. Centromeres specify the kinetochore formation sites on individual chromosomes, and are epigenetically marked by the assembly of nucleosomes containing the centromere-specific histone H3 variant, CENP-A^{3–12}. Although the underlying mechanism is unclear, centromere inheritance is probably dictated by the architecture of the centromeric nucleosome. Here we report the crystal structure of the human centromeric nucleosome containing CENP-A and its cognate α -satellite DNA derivative (147 base pairs). In the human CENP-A nucleosome, the DNA is wrapped around the histone octamer, consisting of two each of histones H2A, H2B, H4 and CENP-A, in a left-handed orientation. However, unlike the canonical H3 nucleosome, only the central 121 base pairs of the DNA are visible. The thirteen base pairs from both ends of the DNA are invisible in the crystal structure, and the α N helix of CENP-A is shorter than that of H3, which is known to be important for the orientation of the DNA ends in the canonical H3 nucleosome¹³. A structural comparison of the CENP-A and H3 nucleosomes revealed that CENP-A contains two extra amino acid residues (Arg 80 and Gly 81) in the loop 1 region, which is completely exposed to the solvent. Mutations of the CENP-A loop 1 residues reduced CENP-A retention at the centromeres in human

cells. Therefore, the CENP-A loop 1 may function in stabilizing the centromeric chromatin containing CENP-A, possibly by providing a binding site for *trans*-acting factors. The structure provides the first atomic-resolution picture of the centromere-specific nucleosome.

Octasome and hemisome models have been proposed for the CENP-A nucleosome architecture¹⁴. In the octasome model, two each of histones H2A, H2B, H4 and CENP-A form a histone octamer, and about 150 base pairs of DNA are wrapped in a left-handed orientation around the histone octamer, as in the canonical H3 nucleosomes^{15–17}. In the hemisome model, however, one each of histones H2A, H2B, H4 and CENP-A^{18,19} form a heterotypic tetramer, and the DNA is wrapped in a right-handed orientation around the tetramer²⁰.

To reveal the architecture of the CENP-A nucleosome, we prepared the nucleosome using bacterially expressed human histones H2A, H2B, H4 and CENP-A^{16,21} (Supplementary Fig. 1) and a 147-base-pair palindromic DNA, which was designed from a human α -satellite sequence containing a binding site for the centromeric protein, CENP-B (Supplementary Fig. 2). The CENP-A nucleosome was reconstituted by a salt-dialysis method, and crystallized as described (Supplementary Methods). The structure was determined at 3.6 Å resolution (Fig. 1, Supplementary Table 1). The crystal structure revealed a histone octamer containing two each of histones H2A, H2B, H4 and CENP-A, with the DNA wrapped in a left-handed orientation around the histone octamer

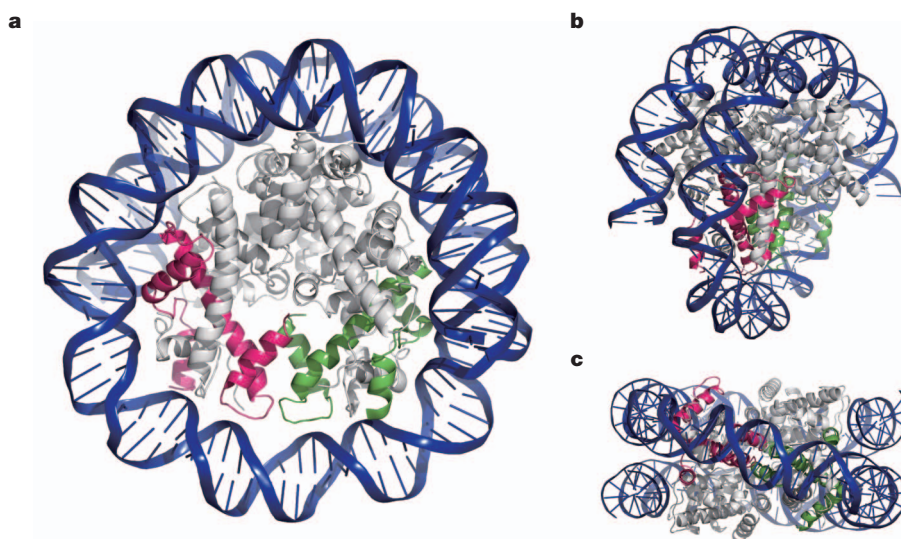


Figure 1 | Crystal structure of the human CENP-A nucleosome. **a–c**, Three views of the CENP-A nucleosome structure are presented. **a**, View in the axis of the DNA supercoil; **b**, **c**, views from the side of the DNA supercoil. Two CENP-

A molecules are shown in magenta and green, respectively. The central 121-base-pair DNA region, which is visible in the CENP-A nucleosome structure, is shown in dark blue.

¹Laboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan. ²Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan. ³Division of Macromolecular Crystallography, Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. ⁴Protein Design Laboratory, Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

*These authors equally contributed to this work.

(Fig. 1). The overall structure is quite similar to those of previously solved nucleosomes containing other histone H3 variants^{13,22,23}. The left-handed DNA wrapping in the crystal structure was also observed in biochemical experiments. CENP-A oligonucleosomes assembled on circular plasmid DNA by a salt-dialysis method or in the presence of human histone chaperones, NAP1 and somatic nuclear autoantigenic sperm protein (sNASP)^{24,25}, under physiological salt conditions, introduce left-handed (negative) supercoils in the DNA (Supplementary Figs 3 and 4). These results strongly indicate that the DNA wrapped in a left-handed orientation is the predominant form in the human CENP-A nucleosome.

In the free CENP-A–H4 structure reported previously²⁶, the CENP-A–CENP-A interface is substantially rotated relative to the H3–H3 interface, indicating that the CENP-A–H4 tetramer may be more compact than the H3–H4 tetramer. However, this specific shape of the CENP-A–CENP-A interface may only be observed in the free CENP-A–H4 tetramer, because the CENP-A–CENP-A interface in the present structure was nearly identical to that of the H3–H3 interface in the canonical H3 nucleosome. Consistently, small-angle X-ray scattering (SAXS) measurements of CENP-A and H3 nucleosomes generated nearly identical SAXS curves and distance distribution functions (Supplementary Fig. 5). These results indicate that the global structures of the CENP-A and H3 nucleosomes are quite similar. In addition, a compact CENP-A nucleosome model containing 147 base-pair DNA constructed from the free H3–H4 tetramer structure²⁶ generated a significantly different SAXS curve and distance distribution function from those of a CENP-A nucleosome model containing 147 base pair DNA based on the CENP-A nucleosome crystal structure, which was very similar to the experimental data (Supplementary Fig. 6). Therefore, the CENP-A nucleosome structure is probably not compacted as was suggested previously²⁶.

Although the overall structure of the CENP-A nucleosome resembles that of the canonical H3 nucleosome, in the CENP-A nucleosome structure, only the central 121 base pairs of the DNA are visible, and thus the thirteen base pairs from both ends of the DNA are disordered (Fig. 1 and Supplementary Fig. 7). This observation indicates that the DNA regions at the entrance and exit of the CENP-A nucleosome lack a fixed conformation. This is consistent with the previous report showing that the DNA segments at the entrance and exit of the CENP-A nucleosome are more flexible than those of the canonical H3 nucleosome^{26–28}. We found that the CENP-A nucleosome can be reconstituted on a 121 base pair DNA (Supplementary Fig. 8). The CENP-A nucleosome induced supercoils into plasmid DNA less efficiently than the H3 nucleosome, which

also indicates that the DNA is partially unwrapped in the CENP-A nucleosomes (Supplementary Fig. 9). Furthermore, the exonuclease assay revealed that the DNA ends of the CENP-A nucleosomes were more susceptible to exonuclease III digestion, compared to those of the H3 nucleosome (Supplementary Fig. 10). Similar results were obtained with other DNA sequences without the CENP-B box (Supplementary Fig. 10). Moreover, the SAXS data showed that the maximum dimension D_{\max} ($D_{\max} = 165 \text{ \AA}$) of the CENP-A nucleosome is slightly longer than that of the H3 nucleosome ($D_{\max} = 147 \text{ \AA}$), which probably reflects the unwrapped conformations of the thirteen base pairs from both ends of the DNA in the CENP-A nucleosome (Supplementary Fig. 5).

This difference in the DNA end structures can be explained by the structural differences between the amino-terminal regions of CENP-A and H3 (Fig. 2a). Previous crystal structures of the canonical H3 nucleosome revealed that the loop segment preceding the α N helix of H3 interacts with the ends of the DNA, and seems to stabilize their orientations (Fig. 2b)^{13,22,23}. Thus, the length of the α N helix seems to be an important factor for maintaining the DNA orientation at the entrance and exit of nucleosomes. The α N helix of CENP-A is at least one helical turn shorter than that of H3, and the preceding region is completely disordered (Fig. 2b). The DNA conformations at the entrance and exit of the nucleosome are clearly related to the organization of the nucleosomes²⁹, particularly within the heterochromatin, where the nucleosomes are presumably tightly packed. A previous study indicated that the centromeric DNA binding protein, CENP-B, binds efficiently to its recognition sequence³⁰, if the sequence is located near the entrance or exit of the CENP-A nucleosome¹⁶. The flexible nature of both ends of the DNA in the CENP-A nucleosome structure may be an inherent property that facilitates the binding of CENP-B, and possibly other centromeric DNA-binding proteins. Thus, the α N helix of CENP-A may have a key role in allowing the DNA to adopt the specific conformations at the entrance and exit of the nucleosome, which are not observed in the canonical H3 nucleosome.

It is notable that CENP-A–DNA contacts may exist at the flexible DNA ends of the CENP-A nucleosome. In the exonuclease digestion experiments (Supplementary Fig. 9), we observed DNA fragments longer than 121 base pairs, suggesting that the CENP-A–DNA contacts may extend beyond 121 base pairs of DNA. Furthermore, the CENP-A nucleosome reconstitution efficiency was slightly lower with a 121-base-pair DNA, compared to that with a 147-base-pair DNA (Supplementary Fig. 8). A possible interpretation is that the CENP-A–DNA contacts extending beyond 121 base pairs of DNA may stabilize

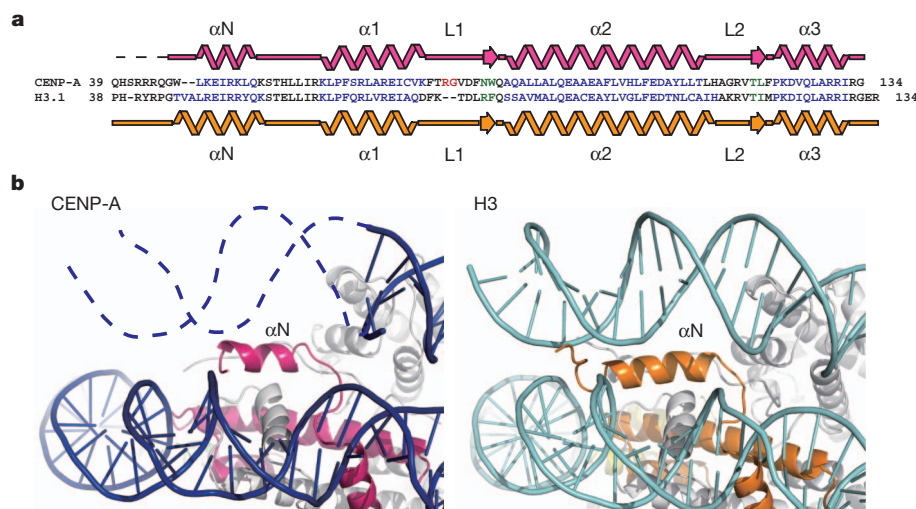


Figure 2 | Structure of the DNA entrance and exit of the human CENP-A nucleosome. **a**, Secondary structure of CENP-A in the nucleosome. The sequences of human CENP-A and H3 are aligned with the secondary structure elements. **b**, Close-up views of the α N helices and the DNA edge regions of the

CENP-A (left panel) and H3 (right panel) nucleosomes. The dashed line in the left panel shows the DNA region that is not visible in the crystal structure. The CENP-A and H3 molecules are shown in magenta and orange, respectively.

the CENP-A nucleosome, resulting in higher reconstitution efficiencies on a 147-base-pair DNA.

A superimposition of CENP-A and H3 reveals a clear difference in the loop 1 region (residues Phe 78–Phe 84 of CENP-A), where CENP-A has two extra amino acid residues (Arg 80 and Gly 81) compared to H3 (Fig. 2a). The CENP-A loop 1 protrudes from the CENP-A nucleosome, and the Arg 80 and Gly 81 residues are located at the tip of the loop (Fig. 3 and Supplementary Fig. 11). In the free CENP-A–H4 tetramer structure, the loop 1 region is more flexible than the other CENP-A regions, as judged from the B-factors. By contrast, in the CENP-A nucleosome, the B-factors of the loop 1 region are similar to those of the other regions (Supplementary Fig. 12), indicating that CENP-A nucleosome formation may stabilize the loop 1 region. The tip of the loop 1 region is solvent-accessible (Fig. 3), and may function as a binding site for *trans*-acting factors that interact directly with the CENP-A nucleosome. To test the functional significance of the CENP-A Arg 80 and Gly 81 residues, we co-expressed the green fluorescent protein (GFP)-tagged CENP-A and the red fluorescent protein (RFP)-tagged CENP-A(del), in which the Arg 80 and Gly 81 residues were deleted, in human-telomerase-immortalized retina pigment epithelial (hTERT-RPE1) cells. Within 1 or 2 days after transfection, both GFP-tagged CENP-A and RFP-tagged CENP-A(del) were recruited to the centromeres, which were identified by a constitutive centromere protein, CENP-C (Fig. 4a). This result indicates that the Arg 80 and Gly 81

residues are not essential for targeting CENP-A to the centromeres. However, 3 days after transfection, the number of cells in which the CENP-A(del) signals were detected at the centromeres was markedly reduced (Fig. 4b). Concomitantly, the number of cells with the CENP-A signal alone increased (Fig. 4b). Similar results were obtained when the fluorescent labels were swapped between CENP-A and CENP-A(del), showing that the phenomenon does not depend on the fusion partner (Fig. 4c). These results indicate that CENP-A(del) is less stably incorporated into centromeres, compared to CENP-A. In addition, two CENP-A mutants, one containing the Arg 80–Gly 81 to Ala 80–Ala 81 substitution (CENP-A(A80A81)) and another with the Val 82–Asp 83 deletion, which disrupts the Arg 80–Gly 81 protrusion (CENP-A(del82–83)), were targeted to centromeres at levels comparable to those of CENP-A, 1 day after transfection. The number of cells retaining the CENP-A mutants at the centromeres also decreased, 3 days after transfection (Supplementary Fig. 13a, b). However, like CENP-A, the CENP-A mutant containing the Val 82–Asp 83 to Ala82–Ala83 substitution (CENP-A(A82A83)) remained at the centromeres, 3 days after transfection (Supplementary Fig. 13c). Thus, the Arg 80 and Gly 81 residues and the size of the protruding loop 1 are critical for stable CENP-A retention at centromeres.

There has been much debate over the CENP-A nucleosome structure and its role in the centromere-specific chromatin structure. Because CENP-A has lower sequence homology to H3, compared to

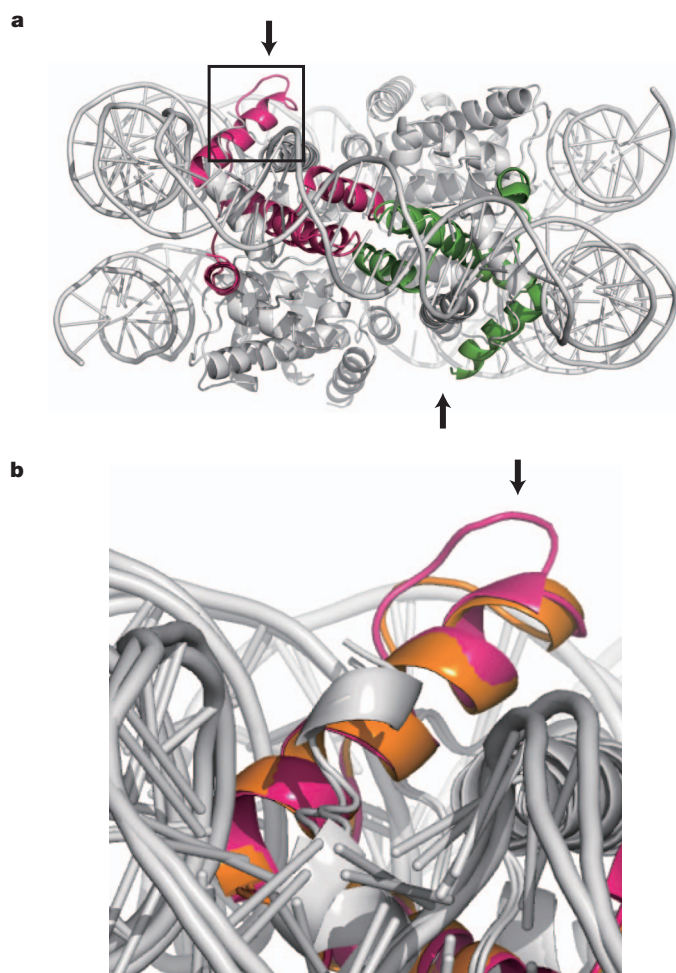


Figure 3 | Structural differences in the loop 1 regions between CENP-A and H3. **a**, Side view of the CENP-A nucleosome. The CENP-A molecules are shown in magenta and green. The box indicates the region enlarged in panel **b**. **b**, Superimposition of the CENP-A (magenta) and H3 (orange) loop 1 regions. Arrows indicate the tip of the CENP-A loop 1 containing the Arg 80 and Gly 81 residues.

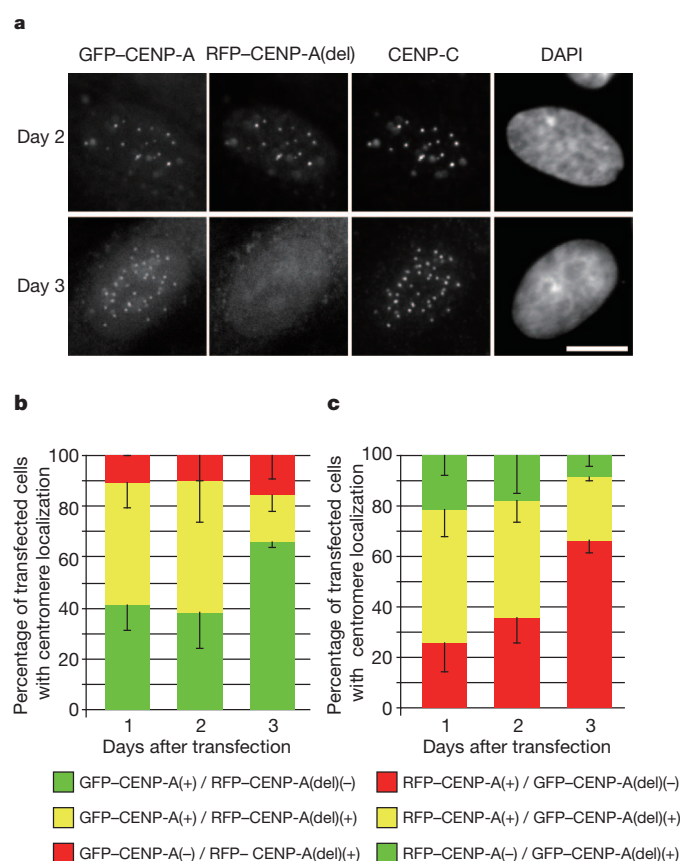


Figure 4 | Less stable association of CENP-A(del) with the centromere.

a, Fluorescence images. hTERT-RPE1 cells were transfected with GFP-tagged CENP-A and RFP-tagged CENP-A(del), fixed, and stained with anti-CENP-C (Cy5) and DAPI. Bar, 10 μ m. **b**, Quantitative data. Using images such as those in panel **a**, the numbers of transfected hTERT-RPE1 cells showing GFP-CENP-A, RFP-CENP-A(del), or both at centromeres were counted ($n > 28$), and the average percentages from three independent transfections were plotted with the standard deviations. **c**, hTERT-RPE1 cells were transfected with GFP-tagged CENP-A(del) and RFP-tagged CENP-A, and were analysed as described in panel **b** ($n > 20$).

other H3 variants, the possibility of a CENP-A nucleosome composed of one of each core histone (hemisome) is an attractive proposition. However, our findings support the octasome model for the CENP-A nucleosome. It is still possible that both types of CENP-A nucleosomes, octasome and hemisome, coexist in the functional centromeric chromatin *in vivo*. We also cannot exclude the possibility that CENP-A hemisomes can be reconstituted under different conditions and/or with factor(s) required for their assembly. Nevertheless, the present structure suggests that the fundamental principles involved in nucleosome formation are likely to be similar among the H3 variants, including CENP-A. The flexibility exclusively observed in the DNA regions located at the entrance and the exit of the CENP-A nucleosome and the loop 1 region protruding from the CENP-A nucleosome may have an essential role in the centromeric chromatin architecture.

METHODS SUMMARY

Human CENP-A, H2A, H2B, H3.1 and H4 were overexpressed in *Escherichia coli* cells, and were purified by a method described previously^{21,23–25}. Details are provided in Methods. The 147-base-pair DNA used in the CENP-A nucleosome reconstitution was prepared by self-ligation with the 71-base-pair fragment of a human α -satellite sequence^{16,21}, containing an extra 5-base overhang, 5'-GTAAC-3', for the cohesive end. The resultant 147-base-pair DNA contained the CENP-B box near both edges, and an A:A mismatch was located at the centre of the DNA (Supplementary Fig. 2). The preparation, crystallization and structural determination of the CENP-A nucleosome are described in Methods. Analyses of the fluorescent protein-tagged CENP-A or CENP-A mutant incorporation at centromeres were performed using hTERT-RPE1 cells. Details are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 August 2010; accepted 1 June 2011.

Published online 10 July 2011.

- Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature Rev. Mol. Cell Biol.* **9**, 33–46 (2008).
- Santaguida, S. & Musacchio, A. The life and miracles of kinetocores. *EMBO J.* **28**, 2511–2531 (2009).
- Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. & Margolis, R. L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).
- Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9**, 573–586 (1995).
- Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell* **94**, 607–613 (1998).
- Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. A histone-H3-like protein in *C. elegans*. *Nature* **401**, 547–548 (1999).
- Henikoff, S., Ahmad, K., Platero, J. S. & van Steensel, B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl Acad. Sci. USA* **97**, 716–721 (2000).
- Howman, E. V. *et al.* Early disruption of centromeric chromatin organization in centromere protein A (*Cenpa*) null mice. *Proc. Natl Acad. Sci. USA* **97**, 1148–1153 (2000).
- Takahashi, K., Chen, E. S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219 (2000).
- Blower, M. D. & Karpen, G. H. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature Cell Biol.* **3**, 730–739 (2001).
- Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. A. Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* **153**, 1209–1226 (2001).
- Régnier, V. *et al.* CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Mol. Cell Biol.* **25**, 3967–3981 (2005).
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Talbert, P. B. & Henikoff, S. Histone variants—ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
- Yoda, K. *et al.* Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 7266–7271 (2000).
- Tanaka, Y. *et al.* Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. *J. Biol. Chem.* **280**, 41609–41618 (2005).
- Camahort, R. *et al.* Cse4 is part of an octameric nucleosome in budding yeast. *Mol. Cell* **35**, 794–805 (2009).
- Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. Tetrameric structure of centromeric nucleosomes in interphase *Drosophila* cells. *PLoS Biol.* **5**, e218 (2007).
- Dalal, Y., Furuyama, T., Vermaak, D. & Henikoff, S. Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl Acad. Sci. USA* **104**, 15974–15981 (2007).
- Furuyama, T. & Henikoff, S. Centromeric nucleosomes induce positive DNA supercoils. *Cell* **138**, 104–113 (2009).
- Tanaka, Y. *et al.* Expression and purification of recombinant human histones. *Methods* **33**, 3–11 (2004).
- Tsunaka, Y., Kajimura, N., Tate, S. & Morikawa, K. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res.* **33**, 3424–3434 (2005).
- Tachiwana, H. *et al.* Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl Acad. Sci. USA* **107**, 10454–10459 (2010).
- Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins *in vitro*. *Nucleic Acids Res.* **36**, 2208–2218 (2008).
- Osakabe, A. *et al.* Nucleosome formation activity of human somatic nuclear autoantigenic sperm protein (sNASP). *J. Biol. Chem.* **285**, 11913–11921 (2010).
- Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A-H4)₂ reveals physical features that mark centromeres. *Nature* **467**, 347–351 (2010).
- Conde e Silva, N. *et al.* CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. *J. Mol. Biol.* **370**, 555–573 (2007).
- Kingston, I. J., Yung, J. S. & Singleton, M. R. Biophysical characterisation of the centromere-specific nucleosome from budding yeast. *J. Biol. Chem.* **286**, 4021–4026 (2011).
- Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. A human centromere antigen (CENP-B) interacts with a short specific sequence in aliphoid DNA, a human centromeric satellite. *J. Cell Biol.* **109**, 1963–1973 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beamline scientists, N. Shimizu, Y. Kawano, M. Makino and T. Hikima, for their assistance with data collection at the BL41XU and BL45XU beamlines of SPring-8. We also thank R. Matsumoto for technical assistance, K. Yoda for anti-CENP-C, and T. Fukagawa and Y. Hiraoka for discussions. This work was supported in part by Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. H.Ku. was also supported by the Waseda Research Institute for Science and Engineering.

Author Contributions H.T., T.S., A.O. and Y.M. purified the histones and CENP-A, crystallized the CENP-A nucleosome, and performed biochemical analyses. H.T., W.K., K.S. and T.S. collected X-ray diffraction data, and H.T., W.K., and S.-Y.P. performed the structural analysis of the CENP-A nucleosome. H.T., A.O., Y.H.-T. and H.Ki. performed the cell biological experiments. T.O., H.T., W.K. and M.S. performed SAXS analysis. H.Ku. conceived, designed and supervised all of the work, and H.Ku., W.K. and H.T. wrote the paper. All of the authors discussed the results and commented on the manuscript.

Author Information The atomic coordinates of the CENP-A nucleosome have been deposited in the Protein Data Bank, under the accession code 3AN2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.Ku. (kurumizaka@waseda.jp).

METHODS

Overexpression of human histones. Human histones H2A and H2B were produced in *Escherichia coli* BL21(DE3) cells, and histone H4 was produced in *E. coli* JM109(DE3) cells. Human CENP-A was produced in *E. coli* DH5 α cells. All histones and CENP-A were produced in *E. coli* cells in the absence of T7 RNA polymerase by omitting the addition of isopropyl- β -D-thiogalactopyranoside, which induces the T7 RNA polymerase production in BL21(DE3) and JM109(DE3) cells. All histones and CENP-A were produced as N-terminal His₆-tagged proteins, as described previously²¹. The His₆ tags of all histones were removed by thrombin protease digestion, leaving a Gly-Ser-His sequence at the N-terminal end of each histone.

For the purpose of structural determination, selenomethionine (Se-Met)-substituted H2B was produced in *E. coli* B834(DE3) cells, using the pET15b vector system (Novagen). The B834(DE3) cells were grown in 100 ml of LB medium for 4 h at 37 °C. The cells were collected and transferred into 300 ml of M9 medium (+50 μ g ml⁻¹ Se-Met). After 12 h growth at 37 °C, the 300-ml culture was added to 2 l of M9 medium (+50 μ g ml⁻¹ Se-Met), and the culture was continued at 37 °C. When the cell density reached 0.5 (*D*₆₀₀), isopropyl- β -D-thiogalactopyranoside (final concentration 1 mM) was added, to induce the expression of H2B. The cells were grown further at 37 °C for 12 h.

Purification of human histones. The cells producing recombinant histones were collected, and were resuspended in 50 ml of buffer A (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM PMSF and 5% glycerol). The cells were disrupted by two rounds of sonication for 200 s each. The cell lysates were centrifuged at 27,216g for 20 min at 4 °C. The supernatants were discarded, and the pellet containing the His₆-tagged histones was resuspended in 50 ml of buffer A, containing 7 M guanidine hydrochloride. The samples were rotated for 12 h at 4 °C, and the supernatants were recovered by centrifugation at 27,216g for 20 min at 4 °C. The supernatants containing the His₆-tagged histones were combined with 4 ml (50% slurry) of nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen), and the samples were rotated for 1 h at 4 °C. The agarose beads were then washed with 100 ml of buffer B (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 6 M urea, 5 mM imidazole, and 5% glycerol). The His₆-tagged histones were eluted by a 100-ml linear gradient of 5 to 500 mM imidazole in buffer B, and the samples were dialysed against buffer C (5 mM Tris-HCl (pH 7.5) and 2 mM 2-mercaptoethanol). The N-terminal His₆ tags were removed from the histones by thrombin protease treatment (1 unit mg⁻¹ of histones; GE Healthcare) at room temperature for 3 h. The removal of the His₆ tags was confirmed by SDS-16% polyacrylamide gel electrophoresis (PAGE); the recombinant histones without the His₆ tag migrated faster than the His₆-tagged histones. After the His₆ tag was uncoupled, each histone was subjected to Mono S column chromatography (GE Healthcare). The column was washed with buffer D (20 mM sodium acetate (pH 5.2), 200 mM NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA, and 6 M urea), and each histone was eluted by a linear gradient of 200 to 800 mM NaCl in buffer D. The purified histones were dialysed against water, and were freeze-dried.

Preparation of DNAs. The 147-base-pair DNA, which was used for reconstituting the CENP-A nucleosome, is a derivative of the human α -satellite DNA (sat4)²¹. The EcoRI site (GAATTC) of the sat4 sequence was replaced by a BstPI site (GGTAACC). The 71-mer DNA fragment containing the 5' half of the sat4 sequence, with the CENP-B box at the edge, was ligated in tandem in the plasmid (p5'Sat4-24). The 71-mer DNA fragment containing an extra 5-base overhang, 5'-GTAAC-3', was prepared for self-ligation according to the method described previously²¹. The 71-mer DNA fragment containing the 5-base overhang was self-ligated, and the palindromic 147-base-pair α -satellite DNA derivative was prepared. The 147-base-pair DNA sequence is: 5'-ATCCTTCGTTGGAAACG GGATTCTTCATTTTCATGCTAGACAGAGAATTCCTCAGTAACCTCTTTG TGCTGGTAACACGACAAAGAAGTTACTGAGAATTCCTCTGTCTAGCAT GAAATGAAGAAATCCCGTTTCCAACGAAGGAT-3'.

In this palindromic 147-base-pair α -satellite DNA derivative, an A:A mismatch was introduced at the centre of the 147-base-pair DNA fragment (underlined).

Preparation of the CENP-A nucleosome. The purified H2A-H2B (Se-Met)-CENP-A-H4 (0.9 mg) and the 147-base-pair DNA (1 mg) were mixed in a solution containing 2 M KCl, and the sample was dialysed against dialysis buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol and 2 M KCl). After dialysis at 4 °C for 3 h, the KCl concentration of the dialysis buffer was gradually decreased to 250 mM with a peristaltic pump (0.8 ml min⁻¹ flow rate). The sample was then dialysed against 10 mM Tris-HCl buffer (pH 7.5), containing 1 mM EDTA, 1 mM dithiothreitol and 250 mM KCl, at 4 °C for 3 h. After this dialysis step, the sample was incubated at 55 °C for 2 h. The CENP-A nucleosome was purified from the free DNA and histones by non-denaturing polyacrylamide gel electrophoresis, using a Prep Cell apparatus (Bio-Rad). The purified CENP-A nucleosome was concentrated, and was dialysed against 20 mM potassium cacodylate buffer (pH 6.0) containing 1 mM EDTA.

Crystallization and structure determination. Crystals of the purified CENP-A nucleosome were obtained by the hanging drop method, after mixing equal volumes of the CENP-A nucleosome solution and 20 mM potassium cacodylate buffer (pH 6.0), containing 60–96 mM KCl and 135–144 mM MnCl₂. The CENP-A nucleosome sample was equilibrated against a reservoir solution of 20 mM potassium cacodylate (pH 6.0), 38–56 mM KCl, and 70–75 mM MnCl₂. Crystals of the CENP-A nucleosome were soaked in a cryoprotectant solution, containing 20 mM potassium cacodylate (pH 6.0), 47 mM KCl, 72 mM MnCl₂, 30% polyethylene glycol 400, and 5% trehalose. The crystals were flash-cooled in a stream of N₂ gas (100 K). The CENP-A nucleosome crystals belonged to the monoclinic space group P2₁, with unit cell constants of $a = 65.8$ Å, $b = 83.3$ Å, $c = 176.8$ Å and $\beta = 100.7^\circ$, and contained one nucleosome in the asymmetric unit. High-resolution diffraction data were obtained using the synchrotron radiation source at the beamline BL41XU station of SPring-8, Harima, Japan.

Diffraction data of the CENP-A nucleosome were integrated and scaled with the HKL2000 program³². The data were processed with the CCP4 program suite³³. The structure was solved by the molecular replacement method, using the MOLREP program³⁴ and the human nucleosome structure (PDB accession number 3AFA) as a search model²³. Most of the amino acid side chains were clearly visible in the map initially calculated at 3.6 Å resolution. Rigid body refinement of the obtained solution was performed using the CNS program³⁵. Further structural refinement consisted of iterative rounds of energy minimization and B factor refinement using the CNS program³⁵, and model building using the COOT program³⁶. The Ramachandran plot of the final structure showed 98.7% of the residues in the most favourable and additional allowed regions, and no residues in the disallowed region. Summaries of the data collection and refinement statistics are provided in Supplementary Table 1. All structure figures were created using the PyMOL program (<http://pymol.org>). The atomic coordinates of the CENP-A nucleosome have been deposited in the Protein Data Bank, with the ID code 3AN2.

Supercoiling assay. Salt-dialysis supercoiling assay. Relaxed plasmid DNA (500 ng) was mixed with 0, 125, 250, 500 and 1,000 ng of histone octamer in 5 μ l of 20 mM Tris-HCl (pH 7.5) buffer, containing 1 mM EDTA, 0.2 mg ml⁻¹ BSA, and 2 M NaCl. The samples were then incubated at 37 °C for 30 min. The NaCl concentration of the sample was reduced to 1 M, 0.8 M, 0.67 M, and 0.2 M by adding dilution buffer, containing 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.2 mg ml⁻¹ BSA, 5 mM MgCl₂, and 0.06 U μ l⁻¹ calf (Invitrogen) or wheat germ (Promega) topoisomerase I. The samples were incubated at 37 °C for 30 min in each dilution step.

Chaperone-mediated supercoiling assay. NAP1 or sNAP (0.25, 0.5, and 1.0 μ M) was pre-incubated with H2A-H2B (150 ng) and CENP-A-H4 (150 ng) at 37 °C for 15 min. Supercoiled plasmid DNA (100 ng), which was relaxed with a topoisomerase I solution (10 mM Tris-HCl (pH 8.0), 2 mM MgCl₂, 5 mM dithiothreitol, and 2 U μ l⁻¹ wheat germ topoisomerase I (Promega)), was added to the reaction mixture. The samples were then incubated at 37 °C for 60 min in 10 mM Tris-HCl (pH 8.0) buffer, containing 140 mM NaCl, 2 mM MgCl₂, and 5 mM dithiothreitol, followed by an incubation at 42 °C for 60 min.

In both the salt-dialysis and chaperone-mediated assays, after the reaction, the samples were treated with 50 μ l of a proteinase K solution (20 mM Tris-HCl (pH 8.0), 20 mM EDTA, 0.5% SDS, and 0.5 mg ml⁻¹ proteinase K (Roche)) at 37 °C for 30 min. The DNA was extracted with phenol/chloroform. The DNA was then precipitated by ethanol, and was analysed by one-dimensional gel electrophoresis on a 1% agarose gel in 1 \times TAE buffer (for the salt-dialysis assay, 1.3 V cm⁻¹ for 15.5 h) or 1 \times TBE buffer (for the chaperone-mediated assay, 1.3 V cm⁻¹ for 15.5 h). For the two-dimensional gel electrophoresis, the DNA was electrophoresed on a 0.7% agarose gel in 1 \times TBE buffer (for the salt-dialysis assay, 2 V cm⁻¹ for 7 h) or a 1% agarose gel in 1 \times TBE buffer (for the chaperone-mediated assay, 1.3 V cm⁻¹ for 15 h) for the first dimension. The gel was then soaked in 1 \times TBE buffer containing 4 mg l⁻¹ of chloroquine for 3 h. The samples were subsequently electrophoresed in 1 \times TBE buffer containing 4 mg l⁻¹ of chloroquine (1.3 V cm⁻¹ for 12 h (for the salt-dialysis assay) or 1.3 V cm⁻¹ for 15 h (for the chaperone-mediated assay)) for the second dimension. The DNA was visualized by SYBR Gold (Invitrogen) staining.

Nucleosome reconstitution by the salt-dialysis method for biochemical analyses. The purified H2A-H2B-CENP-A-H4 or H2A-H2B-H3-H4 octamer was mixed with a DNA fragment (300 μ g, 121-base-pair DNA or 147-base-pair DNA) in a solution containing 2 M KCl (376 μ l). The amounts of histone octamers were 420 μ g for the 121-base-pair DNA and 384 μ g for the 147-base-pair DNA. Nucleosomes were reconstituted and prepared by the same method as described in the 'Preparation of the CENP-A nucleosome' section.

Competitive nucleosome assembly assay. The purified H2A-H2B-CENP-A-H4 octamer (14, 28, 42 or 56 μ g) was incubated in the presence of both the 147-base-pair DNA (24 μ g) and 121-base-pair DNA (20 μ g), in a solution containing 2 M KCl, and the sample was dialysed against dialysis buffer (10 mM Tris-HCl

(pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, and 2 M KCl). After dialysis at 4 °C for 3 h, the KCl concentration of the dialysis buffer was gradually decreased to 250 mM with a peristaltic pump (0.8 ml min⁻¹ flow rate). The sample was then dialysed against 10 mM Tris-HCl buffer (pH 7.5), containing 1 mM EDTA, 1 mM dithiothreitol, and 250 mM KCl, at 4 °C for 3 h. The CENP-A nucleosomes were then analysed by 6% PAGE in 0.2× TBE buffer (18 mM Tris base, 18 mM boric acid, and 0.4 mM EDTA) at 16 V cm⁻¹ for 1 h, followed by ethidium bromide staining.

Nucleosome disruption assay. The CENP-A nucleosomes were reconstituted with a 121-base-pair or 147-base-pair palindromic α -satellite derivative, by the salt dialysis method. The 121-base-pair DNA lacks the 13-base-pair regions from both edges of the 147-base-pair DNA used in the crystallography of the CENP-A nucleosome. The rest of the 121-base-pair DNA sequence is identical to the 147-base-pair palindromic α -satellite derivative. The CENP-A nucleosomes (150 ng) were incubated at 37 °C, 57 °C, 67 °C, 70 °C, or 73 °C for 15 min in the presence of supercoiled plasmid DNA (100 ng). After the incubation, the CENP-A nucleosomes that were not disrupted were separated by non-denaturing 6% PAGE, and were visualized by ethidium bromide staining. The relative band intensities for the CENP-A nucleosomes were quantified and plotted against the temperature.

Exonuclease assay. The reconstituted CENP-A or H3 nucleosomes were treated with 3 units of *Escherichia coli* exonuclease III (Takara), in 10 μ l of 50 mM Tris-HCl (pH 8.0), 5 mM MgCl₂, and 1 mM DTT. After an incubation for 0, 2, 4, or 8 min at 37 °C, the reaction was stopped by the addition of 55 μ l of proteinase K solution (20 mM Tris-HCl (pH 8.0), 20 mM EDTA, 0.5% SDS, and 0.5 mg ml⁻¹ proteinase K (Roche)). After a 15 min incubation at room temperature, the DNA was extracted with phenol/chloroform, precipitated with ethanol, dissolved in Hi-Di Formamide (Applied Biosystems), and then analysed by 10% denaturing PAGE with a gel containing 7 M urea in 0.5× TBE buffer (21 V cm⁻¹ for 1.5 h).

Small-angle X-ray scattering (SAXS). SAXS measurements of the reconstituted CENP-A and H3 nucleosomes, in 20 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 1 mM DTT, were performed at the RIKEN structural biology beam-line I (BL45XU) of SPring-8 (Hyogo, Japan)³⁷. Scattering intensities of the nucleosome solutions were measured with an R-Axis IV⁺⁺ imaging plate detector at 20 °C with a sample-to-detector distance of 3,529 mm, which was calibrated by the powder diffraction from silver docosanoate. Circular averaging of the scattering intensities was then performed to obtain the one-dimensional scattering data $I(q)$ as a function of q ($q = 4\pi\sin\theta/\lambda$), where 2θ is the scattering angle and the X-ray wavelength $\lambda = 0.9$ Å). Three successive measurements were made for each solution, with an exposure time of 60 s. The resultant three data sets were combined after inspections for X-ray radiation damage to the solution and the existence of instrumental artefacts. SAXS measurements of the buffer solution for background subtraction were performed after each measurement of the nucleosome solutions, using the same conditions and procedure as those of the nucleosome solutions. To correct the inter-particle interference effect, $I(q)$ data were collected at four protein concentrations (0.5, 0.7, 1.0 and 1.3 mg ml⁻¹), and extrapolated to zero concentration. The data were processed and analysed using the software applications embedded in the ATSAS package ([http://www.](http://www.embl-hamburg.de/biosaxs/software.html)

[embl-hamburg.de/biosaxs/software.html](http://www.embl-hamburg.de/biosaxs/software.html)). The radius of gyration, R_g , was estimated by fitting the $I(q)$ data using the Guinier approximation $I(q) = I(0) \exp(-q^2 R_g^2/3)$, where $I(0)$ is the forward scattering at the zero scattering angle, in a smaller angle region of $qR_g < 1.3$. Error of R_g was estimated from the least-squares fitting. The distance distribution function $P(r)$ and its error were calculated by the program GNOM³⁸. The maximum dimension D_{max} was estimated from the $P(r)$ function as the distance r , where $P(r) = 0$ (ref. 39), and its error was estimated from the errors of the $P(r)$ values around $P(r) = 0$.

Centromere localization of CENP-A and CENP-A mutants. hTERT-RPE1 cells were transfected with combinations of wild-type CENP-A and CENP-A(Δ del), in which two amino acid residues (the Arg 80 and Gly 81 residues of the CENP-A loop 1) were deleted, and tagged with either GFP or RFP, using GeneJuice (Merck) according to the manufacturer's instructions. hTERT-RPE1 cells were also transfected with combinations of wild-type CENP-A tagged with RFP and CENP-A(Δ del82-83) (where the Val 82 and Asp 83 residues of the CENP-A loop 1 were deleted), or CENP-A(A80A81) (where the Arg 80 and Gly 81 residues were replaced by Ala 80 and Ala 81), or CENP-A(A82A83) (where the Val 82 and Asp 83 residues were replaced by Ala 82 and Ala 83), tagged with GFP. The cells were fixed with 4% paraformaldehyde 1–3 days after transfection, permeabilized, and stained with guinea pig anti-CENP-C⁴⁰ and donkey Cy5-conjugated anti-guinea pig Ig (Jackson ImmunoResearch). DNA was counterstained with 12.5 ng ml⁻¹ DAPI. The fluorescence images were collected using an inverted microscope (Ti-E; Nikon) with a $\times 100$ PlanApo VC numerical aperture (NA) = 1.4 oil-immersion objective lens, or a $\times 40$ PlanApo NA = 0.95 dry lens, equipped with an EM-CCD camera (iXon+; Andor). The numbers of transfected cells exhibiting the GFP- or RFP-tagged protein, or both, at the centromeres were counted, and the average percentages from three independent transfections were plotted with the standard deviations.

- Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2003).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Vagin, A. & Teplyakov, A. *MOLREP*: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Fujisawa, T. *et al.* Small-angle X-ray scattering station at the SPring-8 RIKEN beamline. *J. Appl. Cryst.* **33**, 797–800 (2000).
- Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
- Glatzer, O. & Kratky, O. *Small-angle X-ray Scattering* (Academic Press, 1982).
- Ando, S. *et al.* CENP-A, -B, and -C chromatin complex that contains the I-type α -satellite array constitutes the prekinetochore in HeLa cells. *Mol. Cell. Biol.* **22**, 2229–2241 (2002).

A two-step chemical mechanism for ribosome-catalysed peptide bond formation

David A. Hiller¹, Vipender Singh^{1†}, Minghong Zhong^{1†} & Scott A. Strobel¹

The chemical step of natural protein synthesis, peptide bond formation, is catalysed by the large subunit of the ribosome. Crystal structures have shown that the active site for peptide bond formation is composed entirely of RNA¹. Recent work has focused on how an RNA active site is able to catalyse this fundamental biological reaction at a suitable rate for protein synthesis. On the basis of the absence of important ribosomal functional groups², lack of a dependence on pH³, and the dominant contribution of entropy to catalysis⁴, it has been suggested that the role of the ribosome is limited to bringing the substrates into close proximity. Alternatively, the importance of the 2'-hydroxyl of the peptidyl-transfer RNA⁵ and a Brønsted coefficient near zero⁶ have been taken as evidence that the ribosome coordinates a proton-transfer network. Here we report the transition state of peptide bond formation, based on analysis of the kinetic isotope effect at five positions within the reaction centre of a peptidyl-transfer RNA mimic. Our results indicate that in contrast to the uncatalysed reaction, formation of the tetrahedral intermediate and proton transfer from the nucleophilic nitrogen both occur in the rate-limiting step. Unlike in previous proposals, the reaction is not fully concerted; instead, breakdown of the tetrahedral intermediate occurs in a separate fast step. This suggests that in addition to substrate positioning, the ribosome is contributing to chemical catalysis by changing the rate-limiting transition state.

Several reaction mechanisms have been proposed for peptide bond formation (Fig. 1). The peptidyl transferase reaction occurs through nucleophilic attack of the α -amino group of aminoacyl-tRNA on the carbonyl carbon of the peptidyl-tRNA. A peptide bond forms and the

ester bond linking the peptide to the 3'-oxygen of peptidyl-tRNA breaks, leaving a deacylated tRNA and a peptide lengthened by one amino acid. If substrate positioning is the sole contribution to catalysis, the mechanism within the ribosome is expected to be equivalent to the well-studied uncatalysed reactions, where a pathway involving two tetrahedral intermediates is followed (T^\pm and T^- , Fig. 1; black pathway, see figure legend for explanation of mechanisms and intermediates). At high pH, the rate-limiting transition state for the uncatalysed reaction was predicted⁷ to occur during deprotonation of the zwitterionic T^\pm intermediate, on the basis of the pH-rate dependence of the reaction and a Brønsted coefficient near 1 (transition state D in Fig. 1). At low pH, breakdown of the negatively charged T^- is rate-limiting (transition state F in Fig. 1).

It has also been suggested^{8,9} that the peptidyl-tRNA 2'-hydroxyl acts as a 'proton shuttle' that abstracts a proton from the nucleophilic amine and simultaneously donates another proton either directly or indirectly to the leaving group oxygen. This could happen in a stepwise fashion (Fig. 1, red and orange pathways), involving a tetrahedral intermediate. Alternatively, the reaction has been proposed to be fully concerted, with the amide bond to the nucleophile formed at the same time that the ester bond to the leaving group is broken (Fig. 1, green pathway). Recently, computational methods were used to evaluate concerted proton shuttle mechanisms with and without an additional water molecule, and favoured a fully concerted eight-member proton shuttle (similar to transition state A)¹⁰. Alternatively, other computational studies have indicated a stepwise proton shuttle would be favourable^{11,12}. All of these possibilities have been proposed^{4,8–13}.

Each mechanism predicts a different transition state and hence a different role for the ribosome in catalysis. Therefore we sought to

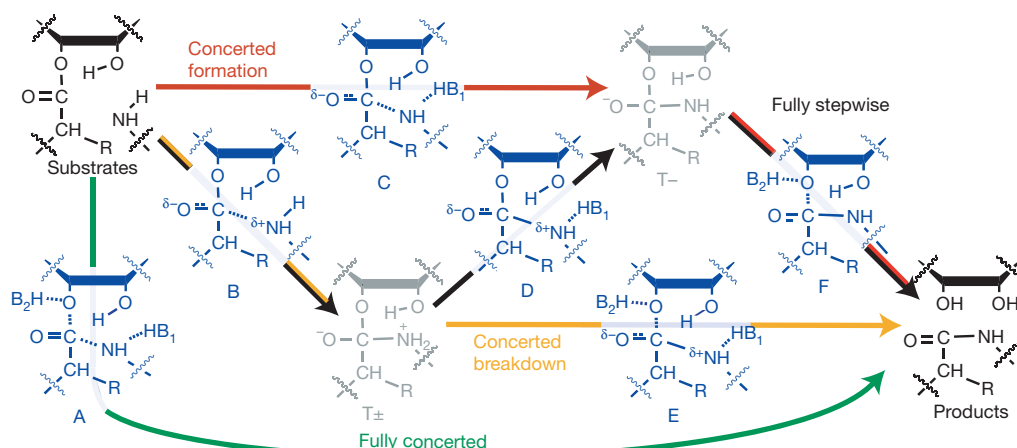


Figure 1 | Proposed reaction mechanisms. Ground states are shown in black, intermediates in grey and transition states in blue. The ribosomal reaction may be fully stepwise with two intermediates, like the uncatalysed reaction (black pathway). Nucleophilic attack leads to an intermediate with a positively charged nitrogen and negatively charged carbonyl oxygen (T^\pm).

Deprotonation leads to a negatively charged intermediate (T^-), which breaks down to products. Alternatively, one of these steps may be concerted, with only one intermediate (red and orange pathways) or the reaction may be fully concerted with no intermediates (green). The identities of B1 and B2 are uncertain; the 2'-hydroxyl may be one, both or neither.

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [†]Present addresses: Department of Chemistry and Bioengineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (V.S.); AMRI, 30 Corporate Circle, Albany, New York 12203, USA (M.Z.).

obtain experimental constraints for modelling the transition state of ribosome-catalysed peptide bond formation using kinetic isotope effect (KIE) analysis (reviewed in ref. 14). Generally, if the rate effect of isotopic substitution (defined as $k_{\text{light}}/k_{\text{heavy}}$) is greater than one (termed 'normal'), it indicates that bonding to that isotope is weaker in the transition state. Reaction coordinate motion of a substituted atom will also contribute to a normal isotope effect. Alternatively, if the isotope effect is less than one (termed 'inverse'), it indicates stronger bonding. The magnitude of the effect is correlated with the change in bond order. To make this analysis possible, we previously synthesized a complete series of molecules that differ by a single isotopic substitution at each atom within the reaction centre^{15,16}. These molecules are substrates for the ribosomal 50S fragment reaction^{17,18}. Previous studies indicate that chemistry is rate-limiting for this assay and that unlike the 70S reaction, which has a complete commitment to catalysis, this reaction is amenable to KIE analysis¹⁸. The Brønsted coefficient of the nucleophile and the structure of the active site are within experimental error for 50S and 70S ribosomes, indicating that the mechanism of catalysis is similar^{6,19}.

To achieve the necessary precision, KIEs were measured by a competitive assay. The light and heavy substrates were incubated in the same reaction, and the change in their ratio as the reaction proceeded was used to determine their relative reaction rates. We used two remote radiolabels, ³²P and ³³P, attached to the 5'-ends of the P-site substrate and performed scintillation counting to define the relative abundance of the two substrates, as previously described for the uncatalysed reaction^{20,21}.

Given the small magnitude of the expected effects (typically in the range of 0.95 to 1.05) and the complexity of the experimental system (a 1.5-MDa complex altered by 1–2 Da), we performed several controls to determine the extent of random and systematic error. First, we demonstrated that the effect from the remote label alone (1.001 ± 0.002 , mean \pm s.e.m.) is much smaller than the effects we expect for atoms at the reaction centre. This value and the effect of deuterium substitution at the amino acid α -carbon were determined 33 and 40 times, respectively. Repeated measurements in both cases yielded normally distributed data, with a significant difference between the mean of the control and the mean of the α -deuterium substitution (Fig. 2). Furthermore, two sets of measurements were made for each substitution, one pairing ³²P with the light isotope at the reaction centre and ³³P with the heavy isotope, and a second with the opposite pairings—that is, ³²P with the heavy isotope and ³³P with the light isotope. These determinations gave the same values within experimental error.

We measured isotope effects for five positions at the reaction centre: the 3'-oxygen leaving group, the carbonyl carbon, the nucleophilic nitrogen, the hydrogen attached to the α -carbon, and the vicinal 2'-hydroxyl (Fig. 3). The simplest effect to interpret is that of ¹⁸O substitution on the leaving group, which measures the extent of C–O bond dissociation in the transition state. Effects in the range of

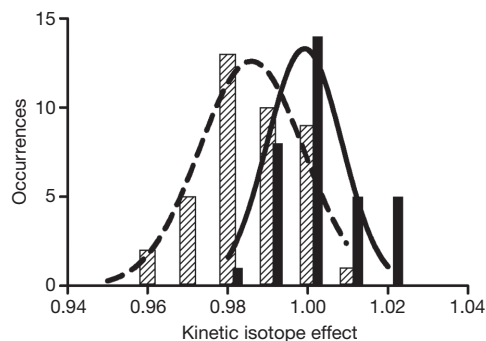


Figure 2 | Histograms of measured isotope effects. Isotope effects on peptide bond formation for controls (solid bars, solid line) and α -deuterium substitution (hatched bars, dashed line) both fit to normal distributions.

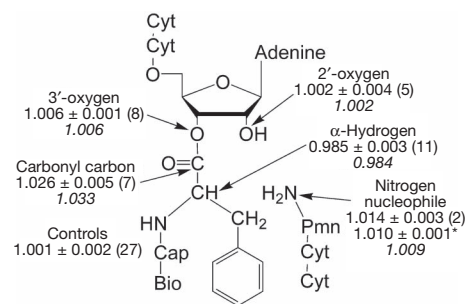


Figure 3 | Kinetic isotope effects. Values are reported in upright font as mean \pm s.e.m., with the number of independent trials in parentheses. The second value for the nitrogen nucleophile (asterisked) was previously determined using mass spectrometry instead of scintillation counting as the readout. Calculated values for the transition state in Fig. 4 are shown in italics. Full substrates are shown: cyt, cytidine; cap, caproic acid; bio, biotin; pmn, puromycin.

1.02 to 1.06 have been observed when cleavage of the C–O bond is rate-limiting^{22,23}. Conversely, if formation of the amide bond or deprotonation of the T \pm intermediate is the rate-limiting step, then a maximum effect of only 1.01 is expected²². We observed a small effect (1.006) on 3'-¹⁸O substitution, which indicates that there is not significant C–O bond-breaking in the rate-limiting step. This is inconsistent with transition state A, the mechanism in which bond formation to the nucleophile is concerted with breaking the bond to the leaving group, and limits the possible transition states to B, C and D (Fig. 1). Notably, the leaving-group oxygen effect previously predicted for fully concerted proton-shuttle mechanisms varied from 1.023 to 1.038 (ref. 10; transition state A), which is inconsistent with the value of 1.006 measured here.

The primary carbonyl-¹³C isotope effect is derived from a combination of bond breaking (that is, the bond to the leaving group oxygen), bond formation (to the nitrogen) and the accompanying change in hybridization. Given that the bond to the leaving group is intact, the last two factors should dominate. The observed effect for the carbonyl carbon is large and normal, 1.026. For hydrazinolysis of methyl formate, a similar effect has been interpreted as resulting from rate-limiting C–N bond formation²⁴. However, it was also shown computationally that an effect of approximately 1.02 could be derived from rate-limiting deprotonation of the zwitterionic intermediate (transition state D)²⁵. The remaining isotope effects made it possible to differentiate between these possibilities.

The deuterium effect on the amino acid α -carbon is primarily derived from the loss of hyperconjugation between the carbonyl π -bond and the antibonding orbital of the C–H bond, which increases the strength of the C–H bond. The magnitude of this effect is largest when the ground state has a large geometric overlap between the π -bond and the antibonding orbital, and the transition state is very close to tetrahedral. Calculations and previous reactions indicate a lower limit of this isotope effect to be approximately 0.96 (refs 21, 26). The measured isotope effect, 0.985, indicates a significant decrease in overlap in the transition state. This is consistent with the decreased π -bond character in a partially tetrahedral transition state.

The KIE of the nucleophilic nitrogen was previously measured to be 1.010, suggestive of an early transition state with partial C–N bond order¹⁸. We remeasured this value using the remote radiolabel-based assay, and again obtained a normal isotope effect, 1.014. Formation of the C–N bond is expected to produce an inverse isotope effect on the nitrogen. The normal isotope effect observed can be derived from two factors: reaction coordinate motion and deprotonation of the nitrogen. Calculations here and elsewhere indicate that either of these factors alone is insufficient to result in a normal isotope effect (Supplementary Tables 1 and 2, and ref. 10). This suggests that the nitrogen is being deprotonated in the rate-limiting step, simultaneous with C–N bond formation (transition state C). Such a mechanism is consistent with the

near-zero Brønsted coefficient⁶, which indicates that there is no buildup of positive charge on the nucleophile in the transition state.

The 2'-hydroxyl contributes 100- to 2,000-fold to the rate of ribosome-catalysed peptide bond formation⁵ and has been postulated to be involved in proton transfer^{8–13}. The isotope effect for this position is close to unity (1.002). Changes in bond order between the 2'-oxygen and protons can be compensated by opposite changes in bond strength to the 2'-carbon; therefore this effect is relatively insensitive to the protonation state of the 2'-oxygen²⁷. The measured effect indicates that the oxygen does not bear a large positive or negative charge, consistent with previous studies¹³.

To supplement the qualitative descriptions above, we calculated isotope effects for several potential transition state structures (Supplementary Tables 1–3). Although we cannot exclude the possibility that multiple steps contribute to the isotope effects, a single transition state structure did yield calculated values in reasonable agreement with the experimental measurements (Fig. 4). As expected for a transition state, there is a single large imaginary frequency corresponding to reaction coordinate motion (carbon–nitrogen bond formation and nitrogen deprotonation). The error inherent in these calculations and our measurements produces an uncertainty in the precise structure of the transition state; however, a single rate-limiting step is consistent with the measured isotope effects. Formation of the tetrahedral intermediate is simultaneous with deprotonation (transition state C). This step is followed by fast breakdown of the tetrahedral intermediate into products (red pathway in Fig. 1). The data are not consistent with a fully concerted reaction mechanism.

The two-step mechanism identified here is markedly different from what has been observed for similar uncatalysed reactions^{7,21}. For uncatalysed reactions at low pH, stepwise formation of T[–] is rate-limiting. At high pH, or with the strong nucleophile hydroxylamine, breakdown of the T[–] intermediate is rate-limiting. Our results indicate that the ribosome not only alters the relative rates of formation and breakdown of tetrahedral intermediates, but also substantially alters the energy landscape such that nucleophilic attack and deprotonation are coordinated.

Concerted attack and deprotonation have also been proposed for aminolysis of the acyl-enzyme intermediate by chymotrypsin^{28,29}. This conclusion was supported in part by a low Brønsted coefficient, similar

to that observed for peptide bond formation. This indicates that the pK_a of the intermediate in the context of these active sites is lower than that of a catalytic group, so that T[±] does not have a finite lifetime and instead T[–] is the only stable intermediate⁷. Although the pK_a of the amine is initially high, it decreases dramatically as the C–N bond is formed. For chymotrypsin, a catalytic histidine is responsible for deprotonating the nucleophile. For the ribosome, the lack of a pH dependence of the reaction argues against general-acid or general-base catalysis³. The 2'-hydroxyl is important for the ribosome reaction, but the pK_a of a hydroxyl dictates that if it abstracts the amino proton it must donate its proton elsewhere. This mechanism, termed the 'proton shuttle', has been proposed in a variety of forms for the ribosome^{8–12}. Alternatively, the 2'-hydroxyl may provide an important transition state hydrogen bond, with another group or a water molecule deprotonating the nucleophile. Either possibility would be consistent with measured solvent isotope effects, which indicate more than one proton is in motion at the transition state (S. Kuhlencoetter and M. Rodnina, personal communication).

The destination of the proton originating from the nucleophile is uncertain. Two computational studies favour the carbonyl oxygen, which would otherwise develop a negative charge^{11,12}. This would result in an uncharged tetrahedral intermediate, which would be consistent with the lack of pH dependence of the reaction. Alternatively, a water molecule observed in crystal structures is correctly positioned to accept a proton from the 2'-hydroxyl and later donate it to the 3'-oxygen leaving group. Finally, the 3'-oxygen could receive the proton from the 2'-hydroxyl. Our data cannot distinguish between these possibilities; for simplicity, we have modelled a water molecule as the proton acceptor.

The ribosome increases the rate of peptide bond formation by an estimated 10⁷-fold through an altered chemical mechanism. The rate of nucleophilic attack is probably increased through positioning effects—first, the increased likelihood of the two substrates being in proximity, and second, precise orientation of the nucleophile (possibly by the 2'-hydroxyl of A2451; ref. 2). The ribosome also has a significant catalytic role, beyond substrate positioning, by coordinating nucleophilic attack and deprotonation in a single rate-limiting step. Catalysis by the ancient, conserved, RNA active site of the ribosome fundamentally alters the reaction pathway for peptide bond formation relative to the uncatalysed reaction.

METHODS SUMMARY

A general strategy for measuring isotope effects with these substrates using a competitive assay has been described²¹. The P-site substrate cytidyl-yl-(3'5')-cytidyl-yl-(3'5')-3'(2')-O-(N-(6-D-(+)-biotinoylamino)hexanoyl)-L-phenylalanyl adenosine (CCApch) was synthesized as described^{15,16}. For each measurement, two substrates differing by a single isotopic substitution at the reaction centre were mixed. Each substrate was labelled with either ³²P or ³³P to allow differentiation by scintillation counting.

Reaction mixtures contained 7 mM MgCl₂, 140 mM NH₄Cl, 25 mM HEPES pH 8.5, trace P-site substrate, 250–500 mM A-site substrate, and 6–10 μM 50S ribosomes. At these concentrations, the isotope effect is on k_{cat}/K_M for the P-site substrate. Aliquots were quenched at 15–30% reacted and near the reaction endpoint. Substrate and product were separated on an acrylamide gel and the fraction reacted determined by phosphorimaging. Each band was eluted and scintillation counted to determine the ratio of ³²P to ³³P in each sample.

To account for substrate that hydrolysed before the reaction was started, the amount of ³²P- and ³³P-labelled product at zero time was subtracted from all determinations of the product isotope ratio. Additionally, some substrate was unreactive even at long time points. The amounts of ³²P- and ³³P-labelled unreactive substrate were similarly subtracted from all determinations of the substrate isotope ratio.

Isotope effects were calculated as previously described²¹, using hybrid density functional methods implemented in Gaussian03. Transition state structures were optimized and frequencies computed on the optimized structures using the three-parameter Becke exchange functional, the LYP correlation functional and the standard 6-31G(d,p) basis set. Isotope effects were calculated from the computed frequencies using ISOEFF 98³⁰. Methylamine was used to model the nucleophile. Calculated isotope effects for tetrahedral intermediate formation were repeated

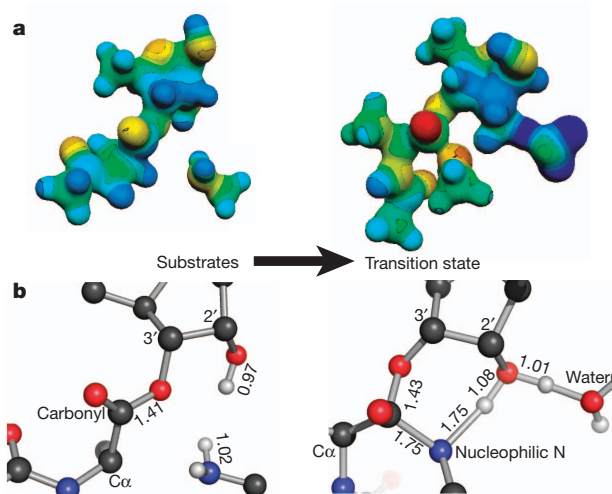


Figure 4 | Structures of the substrates and transition states for peptide bond formation shown in similar orientations. **a**, Electrostatic representations. Blue is electron-deficient (cationic) and red is electron-rich (anionic). **b**, Geometric representations. Carbons are shown grey, oxygens are red, nitrogens are blue, and hydrogens are white. Protons on the nucleophilic nitrogen, 2'-hydroxyl and water are displayed; for clarity, all others are omitted. Bond lengths are shown in Å. The structure was obtained by matching isotope effects calculated with Gaussian03 and Isoeff to the experimental values.

with a set of nucleophiles of increasing size and complexity; these effects varied by less than 0.003.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 April; accepted 3 June 2011.

Published online 17 July 2011.

- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Erlacher, M. D. *et al.* Chemical engineering of the peptidyl transferase center reveals an important role of the 2'-hydroxyl group of A2451. *Nucleic Acids Res.* **33**, 1618–1627 (2005).
- Bieling, P., Beringer, M., Adio, S. & Rodnina, M. V. Peptide bond formation does not involve acid-base catalysis by ribosomal residues. *Nature Struct. Mol. Biol.* **13**, 423–428 (2006).
- Sievers, A., Beringer, M., Rodnina, M. V. & Wolfenden, R. The ribosome as an entropy trap. *Proc. Natl Acad. Sci. USA* **101**, 7897–7901 (2004).
- Zaher, H. S., Shaw, J. J., Strobel, S. A. & Green, R. The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J.* **30**, 2445–2453 (2011).
- Kingery, D. A. *et al.* An uncharged amine in the transition state of the ribosomal peptidyl transfer reaction. *Chem. Biol.* **15**, 493–500 (2008).
- Satterthwait, A. C. & Jencks, W. P. Mechanism of the aminolysis of acetate esters. *J. Am. Chem. Soc.* **96**, 7018–7031 (1974).
- Dorner, S., Panuschka, C., Schmid, W. & Barta, A. Mononucleotide derivatives as ribosomal P-site substrates reveal an important contribution of the 2'-OH to activity. *Nucleic Acids Res.* **31**, 6536–6542 (2003).
- Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell* **20**, 437–448 (2005).
- Wallin, G. & Åqvist, J. The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc. Natl Acad. Sci. USA* **107**, 1888–1893 (2010).
- Wang, Q., Gao, J., Liu, Y. & Liu, C. Validating a new proton shuttle reaction pathway for formation of the peptide bond in ribosomes: a theoretical investigation. *Chem. Phys. Lett.* **501**, 113–117 (2010).
- Rangelov, M. A., Petrova, G. P., Yomtova, V. M. & Vayssilov, G. N. Catalytic role of vicinal OH in ester aminolysis: proton shuttle versus hydrogen bond stabilization. *J. Org. Chem.* **75**, 6782–6792 (2010).
- Huang, K. S., Carrasco, N., Pfund, E. & Strobel, S. A. Transition state chirality and role of the vicinal hydroxyl in the ribosomal peptidyl transferase reaction. *Biochemistry* **47**, 8822–8827 (2008).
- Cleland, W. W. Isotope effects: determination of enzyme transition state structure. *Enzyme Kinetics Mech. D* **249**, 341–373 (1995).
- Zhong, M. & Strobel, S. A. Synthesis of the ribosomal P-site substrate CCA-pcb. *Org. Lett.* **8**, 55–58 (2006).
- Zhong, M. & Strobel, S. A. Synthesis of isotopically labeled P-site substrates for the ribosomal peptidyl transferase reaction. *J. Org. Chem.* **73**, 603–611 (2008).
- Monro, R. E. & Marcker, K. A. Ribosome-catalysed reaction of puromycin with a formylmethionine-containing oligonucleotide. *J. Mol. Biol.* **25**, 347–350 (1967).
- Seila, A. C., Okuda, K., Nunez, S., Seila, A. F. & Strobel, S. A. Kinetic isotope effect analysis of the ribosomal peptidyl transferase reaction. *Biochemistry* **44**, 4018–4027 (2005).
- Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
- Unrau, P. J. & Bartel, D. P. An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc. Natl Acad. Sci. USA* **100**, 15393–15397 (2003).
- Hiller, D. A., Zhong, M., Singh, V. & Strobel, S. A. Transition states of uncatalyzed hydrolysis and aminolysis reactions of a ribosomal P-site substrate determined by kinetic isotope effects. *Biochemistry* **49**, 3868–3878 (2010).
- O'Leary, M. H. & Marlier, J. F. Heavy-atom isotope effects on the alkaline hydrolysis and hydrazinolysis of methyl benzoate. *J. Am. Chem. Soc.* **101**, 3300–3306 (1979).
- Sawyer, C. B. & Kirsch, J. F. Kinetic isotope effects for reactions of methyl formate-methoxyl-¹⁸O. *J. Am. Chem. Soc.* **95**, 7375–7381 (1973).
- Marlier, J. F., Haptonstall, B. A., Johnson, A. J. & Sacksteder, K. A. Heavy-atom isotope effects on the hydrazinolysis of methyl formate. *J. Am. Chem. Soc.* **119**, 8838–8842 (1997).
- Singleton, D. A. & Merrigan, S. R. Resolution of conflicting mechanistic observations in ester aminolysis. A warning on the qualitative prediction of isotope effects for reactive intermediates. *J. Am. Chem. Soc.* **122**, 11035–11036 (2000).
- Hogg, J. L., Rodgers, J., Kovach, I. & Schowen, R. L. Kinetic isotope-effect probes of transition-state structure. Vibrational analysis of model transition states for carbonyl addition. *J. Am. Chem. Soc.* **102**, 79–85 (1980).
- Gawlita, E. *et al.* H-Bonding in alcohols is reflected in the C α –H bond strength: variation of C–D vibrational frequency and fractionation factor. *J. Am. Chem. Soc.* **122**, 11660–11669 (2000).
- Inward, P. W. & Jencks, W. P. The reactivity of nucleophilic reagents with fuorolchymotrypsin. *J. Biol. Chem.* **240**, 1986–1996 (1965).
- Zeeberg, B. & Caplow, M. Transition state charge distribution in reactions of an acetyltyrosylchymotrypsin intermediate. *J. Biol. Chem.* **248**, 5887–5891 (1973).
- Anisimov, V. & Paneth, P. ISOEFF98. A program for studies of isotope effects using Hessian modifications. *J. Math. Chem.* **26**, 75–86 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Klinman and D. Singleton for critical reading of the manuscript, and members of the S.A.S. laboratory for discussion. This work was supported by an NIH postdoctoral fellowship (D.A.H.), a Brown-Coxe fellowship (V.S.) and an NIH grant (GM54839).

Author Contributions D.A.H. and S.A.S. devised experiments, M.Z. synthesized substrates, D.A.H. collected and analysed isotope effect data, V.S. calculated theoretical isotope effects, and D.A.H. and S.A.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A.S. (scott.strobel@yale.edu).

METHODS

Reaction assay. A general strategy for measuring isotope effects with these substrates using a competitive assay has been described²¹. The P-site substrate cytidyl-(3'5')-cytidyl-(3'5')-3'(2')-O-(N-(6-D-(+)-biotinoylamino)hexanoyl)-L-phenylalanyl)adenosine (CCApbc) was synthesized as previously described^{15,16}. Substrates were prepared that differed only by a single isotopic substitution at the reaction centre. Each substrate was 5'-end labelled with either $\gamma^{32}\text{P}$ - or $\gamma^{33}\text{P}$ -ATP by polynucleotide kinase and purified on a pH 6 denaturing 15% polyacrylamide gel. ^{32}P - and ^{33}P -CCApbc were mixed at an approximately 1:4 ratio and purified on a pH 6 non-denaturing 12% polyacrylamide gel. Purified mixes were used as soon as possible to minimize hydrolysis in the starting material.

The isotopic enrichment of each sample was determined by high-resolution Fourier-transform-ion cyclotron resonance mass spectrometry. Each molecule had 95% or greater isotopic purity.

Ribosome reaction mixtures contained 7 mM MgCl_2 , 140 mM NH_4Cl , 25 mM HEPES pH 8.5, trace P-site substrate, 250–500 mM A-site substrate, and 6–10 μM 50S ribosomes. At these concentrations the A-site substrate is saturating but the ribosome concentration is not; therefore the isotope effect is on $k_{\text{cat}}/K_{\text{M}}$ for the P-site substrate. All reactions were performed at room temperature. Reaction aliquots at 15–30% reacted (approximately 1 min) and near the reaction endpoint (greater than 30 min) were quenched with three volumes of formamide loading buffer. Each time point was run on a pH 6 denaturing 15% polyacrylamide gel to separate substrate from product.

Substrate and product bands were visualized using a Storm 840 PhosphorImager with a two-ply sheet of Duck tape between the gel and screen to block ^{33}P emission. The fraction of ^{32}P -labelled substrate reacted could then be determined. The ratio of ^{32}P to ^{33}P was determined by scintillation counting. Each product band was excised from the gel and eluted into 1 ml of 50 mM NaCl overnight. This elution was then added to 13 ml of Optima Gold scintillation fluid and counted for 30 min along with ^{32}P and ^{33}P standards.

Data analysis. Counts per minute were divided into two channels, 0–400 keV and 400–2,000 keV. Approximately 70% of the ^{32}P standard was detected in the high energy channel, and greater than 99% of the ^{33}P sample was detected in the low energy channel. The ratio of ^{32}P to ^{33}P in each sample could be determined using equation (1):

$$^{33}\text{P}/^{32}\text{P} = (A - Br)/(B(1 + r)) \quad (1)$$

where A is the counts per minute in the low energy channel, B is the counts per minute in the high energy channel, and r is the ratio of emission of a ^{32}P standard detected in the low energy channel to the high energy channel. To account for substrate that hydrolysed before the reaction was started, the amount of ^{32}P - and ^{33}P -labelled product at zero time was determined. These were subtracted from all other determinations of the product isotope ratio. Additionally, some substrate was unreactive even at long time points. The amounts of ^{32}P - and ^{33}P -labelled unreactive substrate were similarly subtracted from all other determinations of the substrate isotope ratio. The observed isotope effect was then determined from the ratio in the midpoint and endpoint samples and the fraction reacted using equation (2):

$$\text{KIE} = \log(1 - f)/\log(1 - fR_p/R_0) \quad (2)$$

where f is the fraction reacted, R_p is the isotope ratio in the product at that fraction reacted, and R_0 is the isotope ratio in the product at the reaction endpoint. For

high-fraction-reacted samples (greater than 50%), the isotope effect was also determined from substrate and endpoint samples using equation (3):

$$\text{KIE} = \log(1 - f)/\log[(1 - f)R_s/R_0] \quad (3)$$

where R_s is the isotope ratio in the remaining substrate at f fraction reacted, and R_0 is as above. This value was corrected for incomplete isotopic incorporation as determined by mass spectrometry with equation (4):

$$\text{KIE}_{\text{corrected}} = 1 + (\text{KIE}_{\text{observed}} - 1)/[1 - \text{KIE}_{\text{observed}}(1 - e)] \quad (4)$$

where e is the isotopic enrichment of the heavy sample. This equation assumes a negligible amount of heavy isotope in the light sample^{14,21}.

For the control and α -deuterium isotope effects, enough trials were performed to plot a histogram of the data (Fig. 2). Both data sets were fitted to a normal distribution, equation (5):

$$c = a \exp\{-0.5[(x - \mu)/\sigma]^2\} \quad (5)$$

where c is the number of counts in a given bin, a is the amplitude, μ is the mean and σ is the standard deviation. For some substitutions, the standard errors for trials performed on the same day were smaller than those from different days, indicating that each trial may not be independent. Therefore, for all substitutions each set of experiments from a single day were averaged. The mean and standard error of multiple data sets are reported.

Computation of transition states. Transition state structures that reproduced the experimental KIEs were determined using hybrid density functional methods implemented in Gaussian03. Tetrahydro-4,5-dihydroxy-2-(methoxymethyl)furan-3-yl 2-acetamidopropanoate was used to model the P-site substrate and methylamine was used to model the A-site nucleophile. Calculated isotope effects for tetrahedral intermediate formation were repeated with a set of nucleophiles of increasing size and complexity; the calculated isotope effects varied by less than 0.003.

Structures of the transition states were optimized and the frequencies were computed for the optimized structures using the three-parameter Becke (B3) exchange functional, the LYP correlation functional and the standard 6-31G (d,p) basis set. The 5'-methoxy group and the reaction centre were constrained during the optimization and many of these constraints were modified to match the experimental KIEs.

KIEs and equilibrium isotope effects (EIEs) were calculated from the computed frequencies using ISOEFF 98³⁰. KIEs and EIEs were calculated for a temperature of 298 K and the frequencies were scaled using a factor of 0.964, corresponding to B3LYP/6-31G(d, p). KIEs were calculated whenever the magnitude of the imaginary frequency was greater than 50 $i \text{ cm}^{-1}$, otherwise EIEs were computed. All vibrational modes were used to calculate isotope effects.

Geometric and the electrostatic models were generated by iteratively optimizing the transition states by modifying the applied constraints until the computed isotope effects closely matched the experimental KIEs.

The natural bond orbital (NBO) calculations were performed on optimized structures by including the `pop = nbo` keyword in the route section of input files. The molecular electrostatic potential (MEP) surfaces were calculated by the CUBE subprogram of Gaussian03. The formatted checkpoint files used in the CUBE subprogram were generated by constrained geometry optimization at the B3LYP level of theory with the 6-31G** basis set. MEP surfaces of the substrate and the transition states were visualized using Molekel4.0 at a density of 0.2 electrons per \AA^3 . Geometric figures were created using PyMol.

CORRIGENDUM

doi:10.1038/nature10281

Migrastatin analogues target fascin to block tumour metastasis

Lin Chen, Shengyu Yang, Jean Jakoncic, J. Jillian Zhang
& Xin-Yun Huang

Nature **464**, 1062–1066 (2010)

In this Letter, we reported the crystal structure of macroketone bound to fascin (Protein Data Bank number PDB 3LNA) (Fig. 2d). The chemical structure of macroketone was incorrectly shown and has been corrected (PDB 308K). We have been advised that the crystallographic data for this complex are not technically robust, and do not justify the conclusion that macroketone is bound as shown in Fig. 2d. We therefore regretfully withdraw the X-ray structure models PDB 3LNA and PDB 308K from this study. However, we believe that the rest of the Letter, including the observations made using the mutants, is not directly affected.

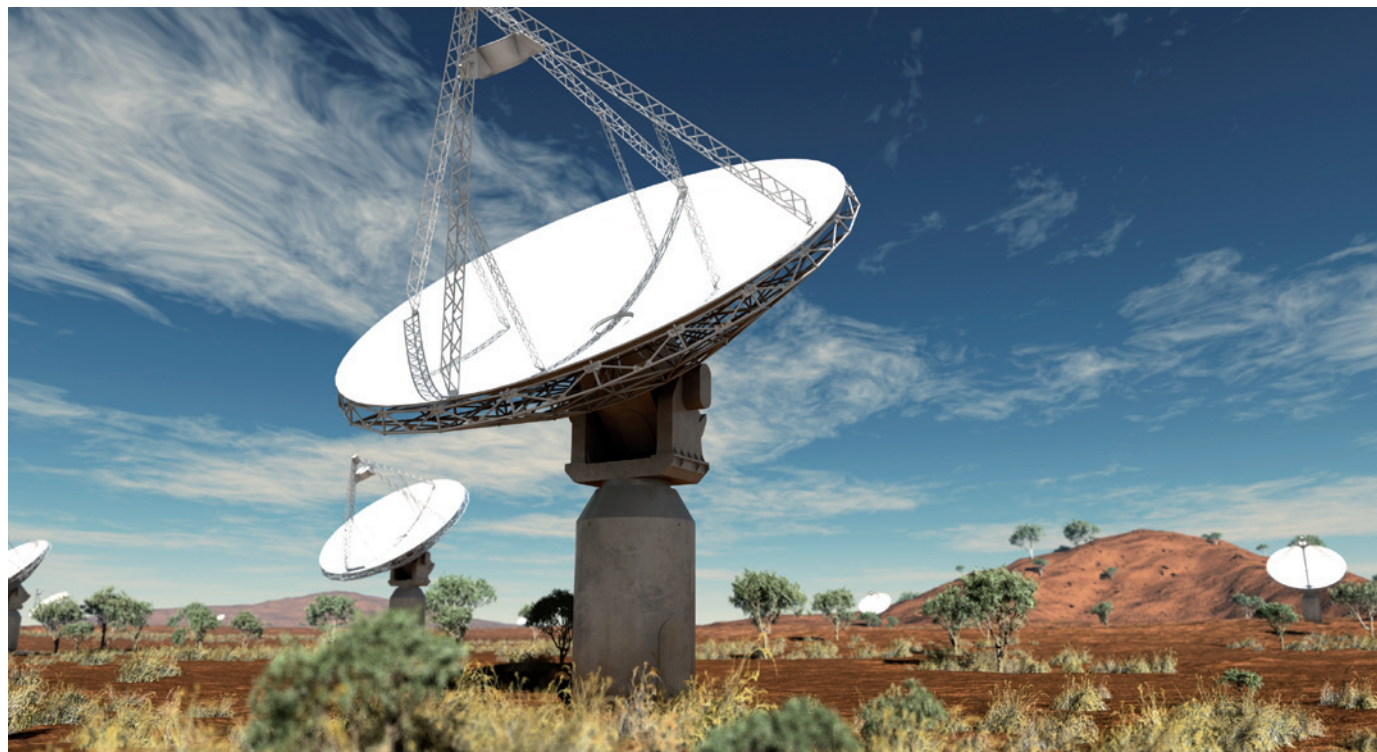
CAREERS

TURNING POINT Chemist's move into energy research earns him a growing lab **p.243**

EUROPEAN UNION PhD scheme will combine science with business innovation **p.243**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

SWINBURNE ASTRONOMY PRODUCTIONS/DESIGN DATA FROM CSIRO



Artist's impression of the Australian Square Kilometre Array Pathfinder radio telescope to be built at the Murchison Radio-astronomy Observatory.

AUSTRALIA

Astronomy in the outback

A bid to host a major international radio telescope has created a demand for astronomers in Western Australia, and built a community that looks set to endure.

BY JAMES MITCHELL CROW

Very little happens in the outback of Western Australia — and that's how a growing group of astronomers likes it. The area is remote from urban centres and industry, so it is extremely radio-quiet: free of the 'noise' generated by cars, mobile phones and other trappings of civilization. That makes it an ideal spot in which to study radio waves from distant corners of the Universe, and is key to the nation's bid to host a prestigious international telescope, the Square Kilometre Array (SKA).

Five years ago, there was not a single astronomy researcher based in the state. Now, with the SKA on the horizon, two advanced pilot radio telescopes planned, high-level government

funding available and a fledgling astronomy community looking to add to its ranks, Western Australia has the potential to be a research hub for years to come. But to succeed, it will have to sustain funding and attract enough researchers to this remote corner of the world.

SKA BAND

The astronomy expansion was sparked in 2006, when an international committee announced that Australia was one of two nations bidding to host the SKA, picked from a shortlist of four. The €1.5-billion (US\$2.1-billion) SKA will be the world's most powerful radio telescope, a continent-spanning array of dishes with a total collecting area of one square kilometre that will combine to make an instrument 50 times

more sensitive than today's best. It is expected to help unlock some of the Universe's biggest mysteries, including the nature of dark energy and dark matter. Next year, the SKA selection panel will make its final decision on whether the telescope will be built on sites spanning Australia and New Zealand, or in South Africa.

Not every astronomer is convinced of the SKA's merits. Some wonder whether scientists and engineers can overcome technological, computing and energy-consumption challenges to maximize the array's effectiveness. "Any project of the scale that we are talking about has people who question whether this is the right way to spend the money," says Anton Zensus, director of the Max Planck Institute for Radio Astronomy in Bonn, Germany. ►



Hundreds of radio antennas could one day populate the outback, as in this artist's rendition.

► The SKA has ample support in Europe, but sentiment is mixed among US astronomers; some suggest that a large optical telescope should be the priority. The array's proponents are pushing ahead anyway, continuing with the site-selection process and developing technologies to enable it. Western Australia's astronomy community has already benefited.

INFRASTRUCTURE BUILD-UP

The University of Western Australia (UWA) and Curtin University, both in Perth, are working hard to bring astronomers to the state. In 2006, the state government awarded two Aus\$1-million (US\$1.1-million) Premier's fellowships to allow the UWA to hire radio astronomers. At the same time, Curtin established a high-profile position in radio astronomy, and another in astronomy engineering.

The academics who filled these posts came up with a plan to expand the local astronomy community. On 1 September 2009, they opened the International Centre for Radio Astronomy Research (ICRAR) in Perth — a joint venture between the UWA and Curtin, established with Aus\$100 million in funding from the universities and the state government to take it through to 2014. Planning for a grant-renewal application to form an expanded ICRAR 2 is already under way, says Peter Quinn, one of the Premier's fellows and the director of ICRAR. The university recruited him from the European Southern Observatory in Munich, Germany.

The centre already employs more than 60 staff members, 55 of whom are researchers in astronomy, engineering or information technology. More than half have been recruited from overseas. It also has 22 PhD students and 20 postdocs. "We aim to be a really strong part of the Australian bid for the SKA," says Quinn, "but we want also to be a major contributor to the international SKA project, independent of where the SKA is placed." The UWA and Curtin have already established seven permanent tenure-track positions at ICRAR.

Alan Duffy is a postdoctoral researcher who joined ICRAR in 2009. He came from one of the world's most prestigious astronomy research institutes, the Jodrell Bank Centre for Astrophysics in Manchester, UK. The culture difference is marked, says Duffy. "It's a lot more vibrant," he adds. "In ICRAR there's a real sense of a bit more adventure going on, as it is so new."

Two key attractions brought Duffy to ICRAR. "In the bigger institutes in Europe, often you're more of a cog in a machine, whereas at ICRAR I had the chance to really define my own research agenda," he says. Duffy runs computer simulations to test theories related to galaxy formation and dark matter.

The other attraction was the infrastructure being built nearby, including two 'pathfinder' radio telescopes that will test some of the technology to be used on the SKA. One is the Aus\$150-million Australian SKA Pathfinder (ASKAP), which will, for example, try out ways of viewing huge areas of the sky at once by combining the signals from multiple antennas. The other is the Murchison Widefield Array, which will consist of 512 antenna elements and will test some of the signal-processing systems to be deployed on the SKA. Both instruments will be located at the Murchison Radio-astronomy Observatory, opened in 2007 about 750 kilometres north of Perth. The observatory will be the base of the SKA if Western Australia's bid is successful.

The raw data from the telescopes will be stored and processed at the Pawsey Centre, an Aus\$80-million supercomputing facility in a



"There's a real sense of a bit more adventure going on, as it is so new."

Alan Duffy

suburb of Perth, which is expected to be among the top 20 fastest supercomputers in the world when it opens in 2013. Astronomy brought the Pawsey Centre to Western Australia, but other scientists will benefit from its presence, including geoscientists and biologists, says Andrew Rohl, executive director of iVEC, the joint venture that will build and run Pawsey.

The investment so far adds up to between Aus\$450 million and \$470 million. This is specially allocated money, notes Quinn, and it has not been taken from existing state or federal government astronomy funds or projects.

IS THE SKA THE LIMIT?

This year, six Australian universities, including the UWA and Curtin, joined together in the Australian Research Council Centre of Excellence for All-Sky Astrophysics (CAASTRO), a consortium aimed at training astronomers. It has Aus\$28 million of federal funding, which goes mainly towards employing postdocs, says Lister Staveley-Smith, deputy director of the consortium. The first recruits have already started work, and the centre will support close to 50 postdocs over the next seven years, with funding renewal possible.

If Australia wins the right to host the SKA, there will probably be another hiring surge. "If the SKA did come to Australia, then I could see ICRAR growing to a major science centre in the 100-plus staff range," says Quinn. Zensus predicts that there will be several hundred researchers affiliated with the SKA at any one time, first commissioning the site and then operating the central array, conducting experiments and running the computer centre.

Zensus notes that Western Australia's workforce and infrastructure are unlikely to "fizzle away", whether the bid is successful or not. The astronomy infrastructure elsewhere in Australia — such as SkyMapper, an optical survey telescope near Canberra — should complement the radio data from ASKAP. "From an academic point of view, there's a critical mass now — you don't notice any isolation because they've brought some of the best people in the world to you," says Duffy. Staveley-Smith agrees. "You wouldn't think that Perth would be the number-one spot for someone from Europe to apply for a job, but we've had some excellent applicants," he says. "I think the astronomy community worldwide recognizes that the government here is being very supportive to astronomy at the moment, coincidentally at a time when things are not quite so good in other places."

"If anything, I'm even more keen to stay in the country than when I first applied," says Duffy, who is a year away from the end of his postdoctoral contract and is applying for an Australian Research Council early-career award. "There's a sense that Australia really is this amazing place to be right now." ■

James Mitchell Crow is a science journalist based in Melbourne, Australia.

TURNING POINT

Joel Rosenthal

Joel Rosenthal, a chemist at the University of Delaware in Newark, was one of 30 young researchers in the United States to win a 2011 Ralph E. Powe Junior Faculty Enhancement Award, which provides research seed money. The award came in April from the Oak Ridge Associated Universities consortium, based in Tennessee.

What was the first turning point in your career?

I always thought I'd be a biologist, but as an undergraduate at New York University (NYU), I discovered a logic behind chemistry — its mechanisms and its quantitative nature — that I found compelling. Using the mechanisms underlying chemical reactions to build compounds and materials was like solving a puzzle. So I started studying organic chemistry. Then I heard Daniel Nocera, an inorganic chemist at the Massachusetts Institute of Technology (MIT) in Cambridge, give a seminar on molecular-level mechanisms of energy conversion. I realized that inorganic chemistry offered many more avenues for researching energy storage and release than did organic chemistry. During Dan's visit to NYU, he encouraged me to apply to MIT for my PhD. I went on to study inorganic chemistry in his lab.

Nocera's work drew media attention. What did that teach you?

Dan has achieved a series of high-profile advances — most recently, an inexpensive solar cell dubbed the artificial leaf. I was in his lab when he first started to attract interest from the mainstream press, including the television news programme *Nightline* and the science show *NOVA*. I began to understand how important it is to communicate chemistry and energy research to the public. Dan's ability to break important problems into small yet meaningful advances was a big part of his success.

What influenced your decision to accept a position at Delaware?

This university and department are under-rated. Delaware is putting lots of resources into equipment such as spectrometers and an advanced X-ray diffractometer. In the current economic climate, this is one of the few institutions that is significantly increasing its research programme and hiring faculty members. There is a great atmosphere in my department, with six young professors all building their research programmes and



collaborating with one another and with senior colleagues.

You assembled a seven-member lab group in your first 18 months at Delaware. Do you worry about over-committing your resources through such fast growth?

It can be difficult to compete with established researchers when recruiting students, but I convinced a postdoc and six first-year graduate students to join me. I'm not worried; in fact, I'm looking to grow further. Society will suffer if scientists don't tackle capture and storage of energy. In my lab, we're trying to generate synthetic fuels by using nickel and palladium electrocatalysts to convert carbon dioxide into, for example, energy-rich carbon monoxide, which is a building block for fuels.

I hope that in the next five to six years, my team and others will have mapped out important molecular-design principles that will help us to design catalysts. I need as many people helping me as possible. The question is whether I bring in enough dollars to support them all. But being here — where costs are much lower than at MIT, for example — has made it easier to quickly establish a group.

What has been the biggest change in energy research in the past few years?

There is a much greater emphasis on applied science, which is important, but I hope that it will not consume the field. There are also more people coming in. We need scientists, but I also see people doing their same old research and wrapping it in an energy context to tap into funding, without trying to be innovative. We need paradigm-shifting fundamental research. ■

INTERVIEW BY VIRGINIA GEWIN

UNITED KINGDOM

More visas for scientists

As part of a policy overhaul, Britain has launched a visa scheme for 'exceptionally talented' people, including scientists. UK immigration policies were revised last year to reduce the number of migrant workers allowed in (see *Nature* 467, 491; 2010); the latest scheme came into effect on 9 August. The science portion will be administered by the Royal Society in London, which can recommend 300 researchers per year whose potential to become leaders in their fields is evidenced by major prizes or fellowships. Applicants need not be sponsored by an employer. Visas are valid for 40 months, after which visiting scientists may apply for a two-year extension, followed by possible settlement.

EUROPEAN UNION

Industrial PhD launched

The European Commission (EC) has created a pilot industrial-PhD programme to unite academia and businesses. The European Industrial Doctorates scheme will have €20 million (US\$28 million) to fund about 100 researchers until 2013, and aims to teach scientists how to combine scientific excellence with business skills. Participating PhD students at partner universities will spend at least half their time at a company in a different European Union (EU) state or associated country. Partner businesses will pay some expenses. EC policy coordinator Vanessa Debais-Sainton says that the programme is likely to become a permanent part of the EU research-funding scheme Horizon 2020, set to launch in 2014.

INNOVATION

Female inventors' prize

A pilot award scheme aims to encourage scientific innovation and invention by women in the European Union (EU). The EU Prize for Women Innovators will award €100,000 (US\$142,000) for first place, €50,000 for second place and €25,000 for third. Applicants must be EU residents, have founded or co-founded a business before 2009 and have received EU funding. Entries will be judged on scientific merit and economic impact. Applications are open until 20 September. European Commission (EC) spokesman Mark English says that the EC hopes to inspire more women to commercialize their research. In future, the award criteria could change to emphasize research at an earlier stage of commercialization.

INTERVENTION

Time out.

BY JOHN GILBEY AND BRIAN MALOW

The marquee outside Building 40 sizzled quietly in the late afternoon Californian sun, a touch of retro flair on a starkly modern building. In front of me in the queue for registration, a famous television cosmologist and a cheerful, bearded philosopher were animatedly comparing their recent flights and the number of time-zones they'd crossed, while the smiling but slightly pressured young organizers dished out name badges, ethically sourced cotton goodie bags and T-shirts emblazoned with the conference logo.

"Hi John! Great to see you again!" Brian, grinning wildly, pumped my hand in welcome. "Bit warmer than London, isn't it? Let's find a drink."

It was cooler inside the building, where the sun was filtered to a murky glow, and as we commandeered a nearby coffee pot I asked Brian if he really thought this was going to work. "Sure it will! You Brits worry way too much. Look, we've narrowed it down to about 20 people — everyone else is too well-known. If there's one here I reckon that these folk have to be our best shot."

I scanned the list. None of the names jumped out as potential conspirators — but I guess that is the whole point. "OK," I sighed resignedly, "I'm game if you are. I suppose if it goes badly wrong we can always say it was a piece of performance art..."

"Good call — you're only 40 miles from San Francisco, remember ... So, heck, anything goes!"

I looked again at the list. "OK, but in that case I get to check out the girls..."

Brian laughed out loud, spilling his coffee, and said something unrepeatable about my lifestyle — then we set out to circulate through the amiable crowd of slightly jet-lagged scientists, philosophers and other vaguely off-beat thinkers.

An hour later, having munched my way along the wonderfully healthy buffet and then quashed any lingering benefits with a couple of artisan beers fresh from the ice bucket, I was no nearer finding whom I was looking for. Everyone was friendly, effusive and seemingly not at all worried by the distinctly odd questioning of a balding, bearded, overweight writer.

Another beer, or two, convinced me that our mission was misdirected, that our premise was flawed and that — in any case — our methodology for delivery was



doomed to ignominy. Most of all, the whole thing suddenly seemed a terrible waste of this beautiful, deep-blue-skied evening in Silicon Valley — surrounded as I was by some of the brightest minds of the age.

It was then that I spotted Brian in animated discussion with a petite, dark-haired girl just next to the ice-cream freezer. Wandering past, I checked her name badge — and was outraged to realize she was on my list.

I have to admit though, that he was very, very good:

"I guess you're here for the same reason I am, yeah?" She smiled at this and made a non-committal move of her head. "Thought so. I mean, the theme of the conference alone is going to attract people like us, am I right? I mean, 'Time as a Flexible Continuum' — there have to be a few of us here, huh...?"

Brian's engaging grin started to thaw her out, and she nodded slowly. "Terrific! I can see we are going to have to swap notes on this. I hadn't realized that they started talking so seriously about temporal stuff this early — I thought it wasn't for another 50 years. I mean, the technical side is way too primitive to bother with — but I'm more into the socio-ethical stuff myself. How about you?

Are you into physics or philosophy?"

She glanced around, spotted me and muttered something

inaudible to Brian. He looked at me, raised one eyebrow a millimetre, offered her his business card and replied casually: "Sure, why not. Oh, when did you say you were from?"

"2158, but I..."

"Quick! Get the net!" he yelled.

Reaching under my jacket, I lurched forward and threw the nylon mist net over the girl. Brian followed up with one of his own and we danced around the crest-fallen, tangled figure, whooping and high-fiving each other in semi-drunken delight.

"Guys! What is this? This is *REALLY* uncool!" A tall, gaunt lady with an organizer's badge stood in front of us with her hands planted firmly on her hips. She did not look happy. From the edge of the growing crowd, two unsmiling security guards started shouldering their way towards us.

"She's from the future! From 2158! She just admitted it!" I yelled at the increasingly restless throng around us. The girl was struggling under the nets, trying to free herself. Some kindly looking folk were attempting to help, with limited success.

"Just think what she knows!" Brian interjected. "With what she tells us we could shortcut maybe a hundred years of research..."

A lean, hard-eyed Californian had joined the organizer lady. "What if she doesn't want to tell you anything?" he asked in a meaningful tone.

"Why, damn it, we'll *MAKE* her!" I yelled across at him.

There was a cry of collective horror — then the people around us became angrily, watchfully quiet. At an unseen signal, the two security staff strode forward and pinioned our arms with professional efficiency.

"Sorry friend, that's not what happens here..."

We were led towards the exit. A knot of familiar, determined-looking science folk moved with us — led by the lean Californian. As this curious circus reached the parking area a van with heavily tinted glass pulled up to the kerb and the side door powered open.

As we were bundled unceremoniously into the van, the Californian turned to us and said quietly: "Did you really think she would risk being here alone?" ■

John and Brian keep bumping into each other at conferences — the rest is fiction, inspired by an original stand-up comedy routine of Brian's.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

Phylogenetic position of *Diania* challengedARISING FROM J. Liu et al. *Nature* **470**, 526–530 (2011)

Liu *et al.*¹ describe a new and remarkable fossil, *Diania cactiformis*. This animal apparently combined the soft trunk of lobopodians (a group including the extant velvet worms in addition to many Palaeozoic genera) with the jointed limbs that typify arthropods. They go on to promote *Diania* as the immediate sister group to the arthropods, and conjecture that sclerotized and jointed limbs may therefore have evolved before articulated trunk tergites in the immediate arthropod stem. The data published by Liu *et al.*¹ do not unambiguously support these conclusions; rather, we believe that *Diania* probably belongs within an unresolved clade or paraphyletic grade of lobopodians.

Without taking issue with the interpretation of *Diania* offered by Liu *et al.*¹, or of the manner in which they coded their characters, we were nonetheless unable to derive their cladogram optimally from the data published. Moreover, we could not replicate their results using any other plausible optimality criteria, or by varying additional parameters not specified by the authors.

Liu *et al.*¹ report analysing their data in PAUP*² under maximum parsimony and with implied weights³ using $k = 2$ (a rather arbitrary choice), but do not mention any other assumptions (for example, the imposition of character order). They obtained three most parsimonious trees, each of 130 steps. Straightforward replication of their stated settings yields 13 trees of just 90 steps each, the strict consensus of which is illustrated (Fig. 1). Why such a difference?

Several of their characters contained inapplicable or gap codings. These appear where a 'daughter' character is logically contingent upon the state of a 'parent', and cannot be coded when the parent is

absent. For example, character 6 (position of frontal appendage) can only be coded in taxa that possess a frontal appendage (character 5) in the first instance (such that a "0" for character 5 necessitates a "-" for character 6). In morphological analyses such as this, inapplicable states are usually assumed to have no bearing on the analysis, being reconstructed passively in the light of known states. In analyses of nucleotide data, by contrast, gaps may alternatively be construed as a fifth and novel state, because shared deletions from some ancestral sequence may actually be informative. If this assumption is made with morphological data, however, all the logically uncodable states in a character are initially assumed to be homologous, and a legitimate basis for recognizing clades. At best, this assigns double weight a priori to absences in the 'parent' character (because the daughter is always contingent), and at worst is positively misleading. This is the approach that we believe Liu *et al.*¹ may have taken. Reanalysis of their data using 'gapmode = newstate' combined with 'collapse = MinBrLen' settings in PAUP*² produced some optimal trees of 130 steps. However, we were still unable to replicate the relationships shown in their Fig. 4, even when varying k between 0 and 10. Rather we either resolved *Diania* in a basal polytomy, or slightly higher in the tree but separated from the arthropods by at least five nodes.

At best, therefore, the position of *Diania* is highly labile and extremely sensitive to the precise methods used. We certainly feel that it is premature to draw conclusions regarding its supposedly pivotal position in the evolution of arthropods. However, our reanalyses do not challenge the more general conclusions of Liu *et al.*¹: namely that the full complement of arthropod characters were probably acquired piecemeal and possibly convergently. Many closely allied groups exploited successfully some but not all of the characters that typify the arthropod crown group. Only in retrospect do we discern a single, ladder-like trajectory through what was really a much more eccentrically branching bush.

Ross C. P. Mounce¹ & Matthew A. Wills¹

¹Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

e-mail: rcpm20@bath.ac.uk

Received 3 March; accepted 24 May 2011.

1. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
2. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4 (Sinauer Associates, 2002).
3. Goloboff, P. A. Estimating character weights during tree search. *Cladistics* **9**, 83–91 (1993).

Author Contributions R.C.P.M. initiated this comment and reanalysed the data. M.A.W. highlighted the potential issue with gap codings. Both authors wrote the note.

Competing financial interests: declared none.

doi:10.1038/nature10266

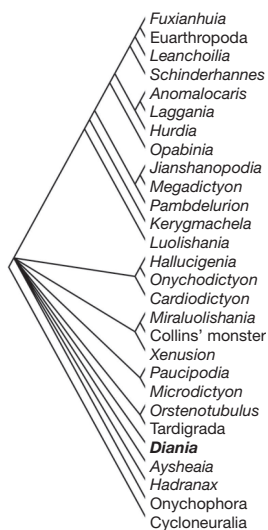


Figure 1 | The strict consensus of 13 most parsimonious trees ($L = 90$) obtained from the published data and settings specified by the authors.

Lobopodian phylogeny reanalysed

ARISING FROM J. Liu *et al.* *Nature* **470**, 526–530 (2011)

Liu *et al.*¹ described an ‘armoured’ lobopodian, *Diania cactiformis*, from the Chengjiang Lagerstätte (China; Cambrian, stage 3); this fossil bears potentially arthropod-like articulated and possibly sclerotized appendages, but lacks a sclerotized body. A cladistic analysis resolved *Diania* as sister-taxon to arthropods. From this phylogenetic position the authors tentatively inferred that arthropodization (sclerotization of limbs) may have preceded arthrodization (sclerotization of body elements) in arthropod evolution. Although we concur with the reasoning behind this inference, it rests on a phylogenetic placement that our analysis of the published data set does not reproduce.

Our analyses were undertaken using implicit enumeration (branch and bound) in TNT (Tree Analysis using New Technology) v.1.1 (ref. 2), first repeating the analysis of Liu *et al.*¹ with implied character weighting ($k = 2$) and additionally with equal character weighting; results are shown in Fig. 1b, c. In neither analysis was *Diania* resolved as sister-taxon to Arthropoda, instead belonging to an unresolved polytomy at the base of the resultant most-parsimonious trees. To confirm that the trees of Liu *et al.*¹ are less parsimonious, a second pair of analyses was undertaken. Here tree topologies were constrained to the published strict-consensus topology (Fig. 1a), again using implied ($k = 2$) and equal weighting schemes. In both cases a single tree was found, substantially longer than those our first analyses recovered (16.4 as opposed to 12.4 steps under implied weighting; 119 as opposed to 89 steps under equal weighting). This demonstrates that the published topology is far from the most parsimonious that can be obtained from the data set. The use of a branch and bound (rather than heuristic) search indicates that this discrepancy is not a result of

chance failure to hit upon the best solution; instead we believe that it reflects a methodological error in the original analysis.

The phylogeny recovered here suggests a polyphyletic origin of arthropodized limbs, where the arthropodization in *Diania* was acquired separately to that of the crown-group arthropods. It does not rule out the ‘leg-first’ (as opposed to arthrodization-first) scenario posited by Liu *et al.*¹, but provides no support for it either. The recovered polychotomy demonstrates that the character coverage of Liu *et al.*¹ is inadequate to resolve the interrelationships of lobopodians; a more comprehensive data set is required to assess the position and significance of *Diania* properly. We note, however, that our results from the Liu *et al.*¹ data set (Fig. 1b) place the euarthropods within the dinocaridids (radiodonts and related taxa), and as such are similar to other analyses^{3,4} which have used more character- and taxon-rich data sets. The origin of arthropodized trunk limbs is problematical in this topology, as dinocaridids apparently lack any form of leg-like trunk appendage⁵. The increasingly detailed fossil record of stem-group euarthropods provides our best chance of resolving this issue, but as yet has failed to do so; unequivocal evidence for any particular ordering of acquisition in these characters is not yet available. *Diania* is a fascinating animal, but in its revised position it contributes little to this debate.

David A. Legg^{1,2}, Xiaoya Ma², Joanna M. Wolfe³,
Javier Ortega-Hernández⁴, Gregory D. Edgecombe² & Mark D. Sutton¹

¹Department of Earth Science and Engineering, Imperial College London, London SW7 2AZ UK.

e-mail: d.legg10@imperial.ac.uk

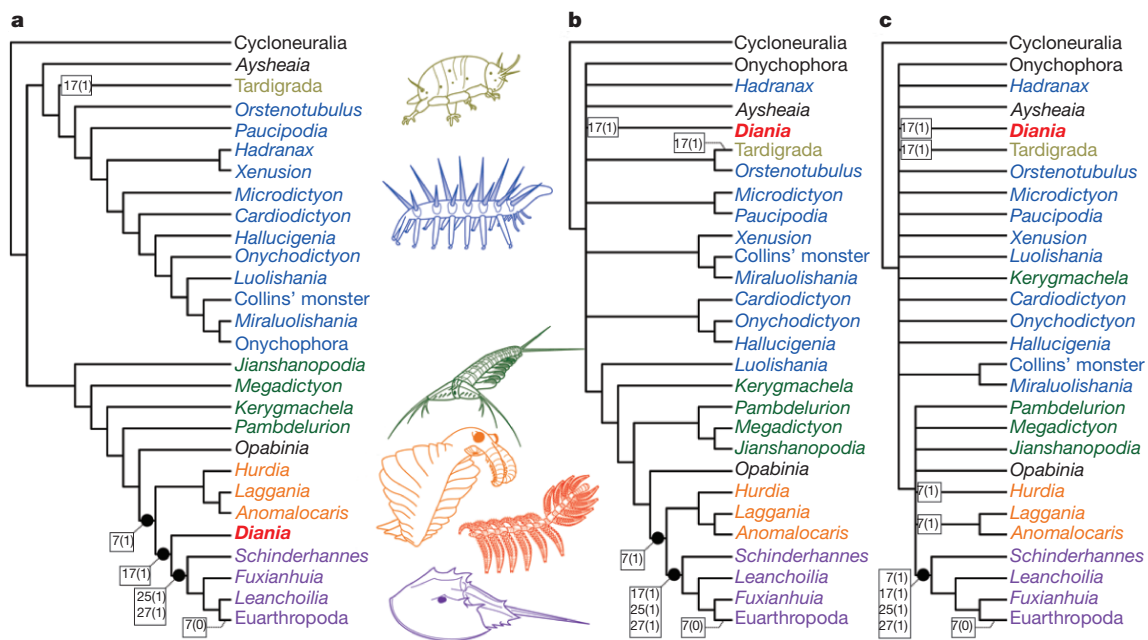


Figure 1 | Phylogeny of Cambrian lobopodians and stem-group arthropods. **a**, Topology proposed by Liu *et al.*¹ (implied weights, $k = 2$, 16.4 steps, consistency index (CI) = 0.336, retention index (RI) = 0.546; equal weights, 119 steps, CI = 0.336, RI = 0.546). **b**, Strict consensus of five most parsimonious trees using the methods of Liu *et al.*¹ (implied weights, $k = 2$, 12.4

steps, CI = 0.444, RI = 0.713). **c**, Strict consensus of 86 trees produced with equal character weighting (89 steps, CI = 0.499, RI = 0.718). Numbers represent important characters of Liu *et al.*¹ relating to arthropodization and arthrodization.

²Department of Palaeontology, Natural History Museum, London SW7 5BD, UK.

³Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA.

⁴Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK.

Received 3 March; accepted 24 May 2011.

1. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
2. Goloboff, P. *et al.* TNT, a free program for cladistic analysis. *Cladistics* **24**, 774–786 (2008).

3. Daley, A. C. *et al.* The Burgess Shale anomalocaridid *Hurdia* and its significance for early euarthropod evolution. *Science* **323**, 1597–1600 (2009).
4. Ma, X. *et al.* Morphology of *Luolishania longicruris* (Lower Cambrian, Chengjiang Lagerstätte, SW China) and the phylogenetic relationships within lobopodians. *Arth. Struct. Dev.* **38**, 271–291 (2009).
5. Zhang, X. & Briggs, D. E. G. The nature and significance of the appendages of *Opabinia* from the Middle Cambrian Burgess Shale. *Lethaia* **40**, 161–173 (2007).

Author Contributions: D.L. was responsible for writing and coordination of the manuscript; X.M., J.W., J.O.-H., G.E.D. and M.S. helped with discussion. D.L. and J.W. were also responsible for reanalysing the data. J.O.-H. produced the accompanying figure and M.S. aided with wording and stylistic editing of the original manuscript.

Competing financial interests: declared none.

doi:10.1038/nature10267

Liu *et al.* reply

REPLYING TO R. C. P. Mounce & M. Wills *Nature* **476**, doi:10.1038/nature10266 (2011); D. A. Legg *et al.* *Nature* **476**, doi:10.1038/nature10267 (2011);

We welcome the reanalyses by Mounce and Wills¹ and Legg *et al.*² of our paper³, and although we do not fully concur with their conclusions we are pleased that *Diania* has reopened the debate about key stages in arthropod evolution. We accept that the position of this fossil remains sensitive to parameters of analysis and in the original publication we conceded that our best-supported tree—*Diania* as sister-group to (*Schinderhannes* + Euarthropoda)—could be subject to

change, and that the ‘walking cactus’ may have a more basal position within the overall framework of the arthropod stem-group. These alternative treatments of our data would seem to confirm this suspicion, although we find the placement of *Diania* in an unresolved, and extremely basal, polytomy alongside velvet worms, tardigrades and various other lobopodians similarly problematical. We do not doubt that the authors’ results^{1,2} are statistically well supported, but what do these cladograms tell us about the evolution of the group? Lobopodians are, by their nature, fairly simple and consequently yield few convincing synapomorphies, either with each other or with arthropods in general. As we discovered, this makes scoring a robust data matrix including both lobopodians and arthropods challenging, and we wonder whether the basal polytomies recovered here are simply due to clustering among taxa with few unequivocal apomorphies and/or much missing data.

Our original placement of *Diania* close to the euarthropods was strongly influenced by the character of jointed trunk appendages. We acknowledge an error in scoring characters 17 and 27 of Tardigrada, which should have been 0 and 1, not 1 and 0, respectively. We also accept the criticism by Mounce and Wills¹ that parent–daughter characters are, to some extent, dependent upon one another and can artificially inflate support for particular clades. This situation is hard to avoid when selecting characters across a range of lobopod/arthropod fossils, which include taxa with unusual morphologies (for example, *Opabinia*) or where there are alternative hypotheses for the homology of a given lobopod feature and its probable arthropod equivalent: should a lobopod/dinocaridid flap be scored the same way as a euarthropod limb

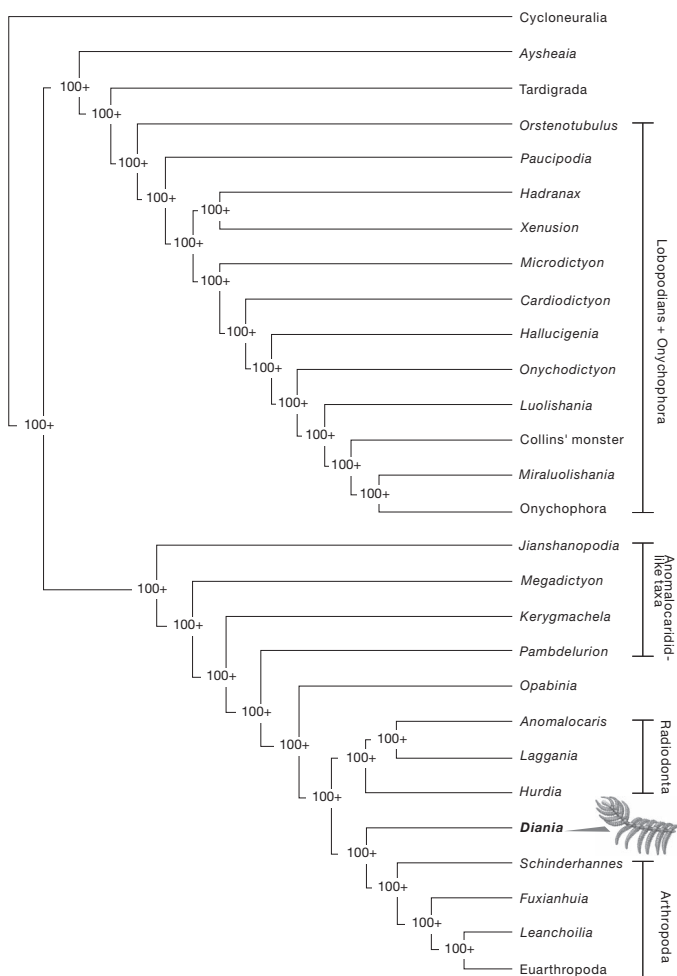


Figure 1 | Reanalysis of our data using PAUP: bootstrap 50% majority-rule consensus tree. Hierarchical structure in the data was assessed using the PTP test⁶ as implemented in PAUP. Parsimonious trees were found through a heuristic search strategy using tree bisection-reconnection (TBR) branch swapping. In an initial analysis, all characters were treated as unordered. Branch support was based on calculating bootstrap values (10,000 replicates and with rearrangement limit of 10,000,000 rearrangements per additional sequence). A significant value in the PTP test ($P = 0.01$) does suggest the presence of a phylogenetic signal in the morphological data, supporting the given topology. Thirty-eight characters were parsimony-informative. The heuristic search for a maximum parsimony solution resulted in only one parsimonious tree (length = 121, homoplasy index (HI) = 0.67, consistency index (CI) = 0.33, retention index (RI) = 0.55, excluding 1 uninformative character). This tree contains 26 resolved nodes, of which many were strongly supported by bootstrap values (100+) and in some nodes bootstrap values showed even greater resolution.

exite; could both be subsumed under a 'biramous limb', etc.? Furthermore, a disadvantage of PAUP is that 'gaps' are automatically treated as missing in the analysis, and the only way to distinguish between them is to treat the gap as a new state; although a gap in our analysis refers to an inapplicable character, which is itself not a new state.

In our initial analysis data were entered into a matrix using MacClade version 3.05 (ref. 4). Analyses were performed with PAUP version 4.0b10 (ref. 5), whereby multi-state characters were treated as 'unordered', other characters were treated as 'ordered', and a branch and bound search under implied weights ($k = 2$) produced many most parsimonious trees—including trees similar to those of Legg *et al.*² and Mounce and Wills¹. For the purpose of character state mapping, we selected one of these most parsimonious trees (see Fig. 4 in ref. 3) which we felt best reflected the complex relationships among lobopodians and arthropods. Also, a constrained analysis was conducted, and three monophyletic groups (lobopodians + Onychophora, Radiodonta, Arthropoda) were enforced. To verify the stability of this tree, we conducted bootstrap analysis and the result (Fig. 1) shows that the contained 26 resolved nodes of this tree were also strongly supported by bootstrap values (100+) and in some nodes bootstrap values

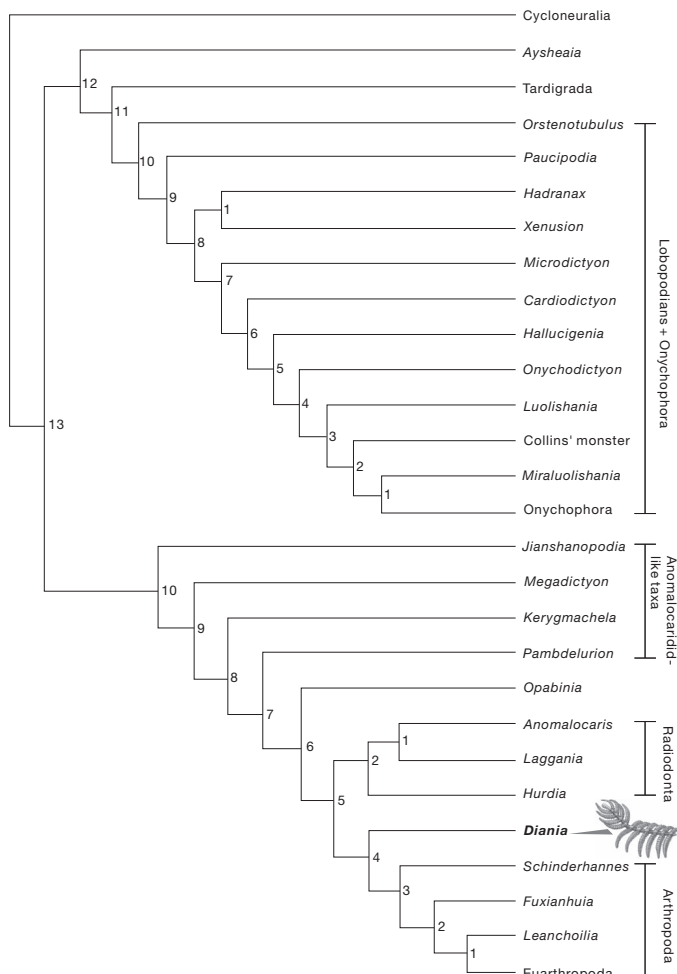


Figure 2 | Tree shows the Bremer support value, which strongly supports *Diania* being sister group to Arthropoda (Bremer index = 4). Four additional steps would be required to disrupt the *Diania*–arthropod relationship. As we noted in the original description, *Diania* is the most arthropod-like of the lobopodians recorded so far—at least in terms of limb morphology—and we feel that it is better placed on the arthropod stem, rather than in an unresolved polytomy.

showed more resolution. Additionally, a significant value of the partitioning tail permutation (PTP) test ($P = 0.01$) suggests the presence of a clear phylogenetic signal in the morphological data, also strongly supporting the topology shown. Furthermore, a test of Bremer support still supports our initial analysis (Fig. 2).

Legg *et al.*² drew attention to the absence of sclerotized trunk limbs in dinocaridids (*Anomalocaris*, etc.), which remains for us one of the great puzzles of stem-group arthropod evolution. Put simply, dinocaridids have a more arthropod-like head region (cephalisation, eyes, sclerotized mouthparts), whereas *Diania* lacks such sophistication in the anterior body region, but has jointed trunk appendages. One solution to this puzzle would be to assume that dinocaridids also had trunk limbs, but lost them secondarily. In this scenario *Diania* could sit comfortably as sister group to a (Dinocaridida + Arthropoda) clade: all three sharing jointed appendages, the latter two sharing cephalisation, etc.

We accept that the tree of Legg *et al.*² and Mounce and Wills¹ is the strict consensus tree, but we feel that the strict tree is, in this case, too conservative to provide meaningful information about the early evolution of the arthropods. This was undoubtedly a complex process, and may have involved numerous parallel developments introducing homoplasy. Under these circumstances, the comb-like most parsimonious tree is unhelpful and exemplifies a disadvantage of parsimony in this instance when faced with a complex early radiation. Indeed, Legg *et al.*² recognized that the strict consensus tree(s) “suggests a polyphyletic origin of arthropodized limbs”. If true, this would be significant. In this context, Mounce and Wills¹ seem to have overlooked the potential significance of their reanalysis of our data. Much evidence has been accumulated that Arthropoda is monophyletic, but the logical conclusion of both reanalyses is that jointed legs are homoplastic, at least within the lobopodian–arthropod assemblage. The implications of this are not trivial, bringing us close to the (largely discredited) Manton school of thought in which jointed appendages were proposed to have evolved in parallel in different (here euarthropod) lineages. At a fundamental level we need to know how easy it is to turn a soft, lobopodian limb into a sclerotized arthropod one, and whether this happened once, twice or on multiple occasions. For this reason alone, we believe that *Diania* is not merely fascinating, but remains invaluable to the evolutionary debate by challenging our notion of what it means to be an arthropod.

Jianni Liu^{1,2}, Michael Steiner², Jason A. Dunlop³, Helmut Keupp², Degan Shu^{1,4}, Qiang Ou⁴, Jian Han¹, Zhifei Zhang¹ & Xingliang Zhang¹

¹Early Life Institute, State Key Laboratory of Continental Dynamics, Department of Geology, Northwest University, Xi'an 710069, China.

²Department of Earth Science, Freie Universität Berlin, D-12249 Berlin, Germany.

e-mail: liu2009@zedat.fu-berlin.de or liujianni@126.com

³Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University Berlin, D-10115 Berlin, Germany.

⁴School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China.

1. Mounce, R. C. P. & Wills, M. Phylogeny of *Diania* challenged. *Nature* **476**, doi:10.1038/nature10266 (2011).
2. Legg, D. A. *et al.* Lobopodian phylogeny reanalysed. *Nature* **476**, doi:10.1038/nature10267 (2011).
3. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
4. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution* Version 3 (Sinauer Associates, 1992).
5. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4.0. (Sinauer Associates, 2002).
6. Faith, D. P. & Cranston, P. S. Probability, parsimony, and Popper. *Syst. Biol.* **41**, 252–257 (1992).

doi:10.1038/nature10268

Phylogenetic position of *Diania* challengedARISING FROM J. Liu et al. *Nature* **470**, 526–530 (2011)

Liu *et al.*¹ describe a new and remarkable fossil, *Diania cactiformis*. This animal apparently combined the soft trunk of lobopodians (a group including the extant velvet worms in addition to many Palaeozoic genera) with the jointed limbs that typify arthropods. They go on to promote *Diania* as the immediate sister group to the arthropods, and conjecture that sclerotized and jointed limbs may therefore have evolved before articulated trunk tergites in the immediate arthropod stem. The data published by Liu *et al.*¹ do not unambiguously support these conclusions; rather, we believe that *Diania* probably belongs within an unresolved clade or paraphyletic grade of lobopodians.

Without taking issue with the interpretation of *Diania* offered by Liu *et al.*¹, or of the manner in which they coded their characters, we were nonetheless unable to derive their cladogram optimally from the data published. Moreover, we could not replicate their results using any other plausible optimality criteria, or by varying additional parameters not specified by the authors.

Liu *et al.*¹ report analysing their data in PAUP*² under maximum parsimony and with implied weights³ using $k = 2$ (a rather arbitrary choice), but do not mention any other assumptions (for example, the imposition of character order). They obtained three most parsimonious trees, each of 130 steps. Straightforward replication of their stated settings yields 13 trees of just 90 steps each, the strict consensus of which is illustrated (Fig. 1). Why such a difference?

Several of their characters contained inapplicable or gap codings. These appear where a ‘daughter’ character is logically contingent upon the state of a ‘parent’, and cannot be coded when the parent is

absent. For example, character 6 (position of frontal appendage) can only be coded in taxa that possess a frontal appendage (character 5) in the first instance (such that a “0” for character 5 necessitates a “-” for character 6). In morphological analyses such as this, inapplicable states are usually assumed to have no bearing on the analysis, being reconstructed passively in the light of known states. In analyses of nucleotide data, by contrast, gaps may alternatively be construed as a fifth and novel state, because shared deletions from some ancestral sequence may actually be informative. If this assumption is made with morphological data, however, all the logically uncodable states in a character are initially assumed to be homologous, and a legitimate basis for recognizing clades. At best, this assigns double weight a priori to absences in the ‘parent’ character (because the daughter is always contingent), and at worst is positively misleading. This is the approach that we believe Liu *et al.*¹ may have taken. Reanalysis of their data using ‘gapmode = newstate’ combined with ‘collapse = MinBrLen’ settings in PAUP*² produced some optimal trees of 130 steps. However, we were still unable to replicate the relationships shown in their Fig. 4, even when varying k between 0 and 10. Rather we either resolved *Diania* in a basal polytomy, or slightly higher in the tree but separated from the arthropods by at least five nodes.

At best, therefore, the position of *Diania* is highly labile and extremely sensitive to the precise methods used. We certainly feel that it is premature to draw conclusions regarding its supposedly pivotal position in the evolution of arthropods. However, our reanalyses do not challenge the more general conclusions of Liu *et al.*¹: namely that the full complement of arthropod characters were probably acquired piecemeal and possibly convergently. Many closely allied groups exploited successfully some but not all of the characters that typify the arthropod crown group. Only in retrospect do we discern a single, ladder-like trajectory through what was really a much more eccentrically branching bush.

Ross C. P. Mounce¹ & Matthew A. Wills¹

¹Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

e-mail: rcpm20@bath.ac.uk

Received 3 March; accepted 24 May 2011.

1. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
2. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4 (Sinauer Associates, 2002).
3. Goloboff, P. A. Estimating character weights during tree search. *Cladistics* **9**, 83–91 (1993).

Author Contributions R.C.P.M. initiated this comment and reanalysed the data. M.A.W. highlighted the potential issue with gap codings. Both authors wrote the note.

Competing financial interests: declared none.

doi:10.1038/nature10266

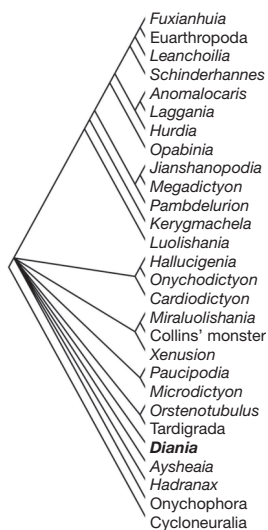


Figure 1 | The strict consensus of 13 most parsimonious trees ($L = 90$) obtained from the published data and settings specified by the authors.

Lobopodian phylogeny reanalysed

ARISING FROM J. Liu *et al.* *Nature* **470**, 526–530 (2011)

Liu *et al.*¹ described an ‘armoured’ lobopodian, *Diania cactiformis*, from the Chengjiang Lagerstätte (China; Cambrian, stage 3); this fossil bears potentially arthropod-like articulated and possibly sclerotized appendages, but lacks a sclerotized body. A cladistic analysis resolved *Diania* as sister-taxon to arthropods. From this phylogenetic position the authors tentatively inferred that arthropodization (sclerotization of limbs) may have preceded arthrodization (sclerotization of body elements) in arthropod evolution. Although we concur with the reasoning behind this inference, it rests on a phylogenetic placement that our analysis of the published data set does not reproduce.

Our analyses were undertaken using implicit enumeration (branch and bound) in TNT (Tree Analysis using New Technology) v.1.1 (ref. 2), first repeating the analysis of Liu *et al.*¹ with implied character weighting ($k = 2$) and additionally with equal character weighting; results are shown in Fig. 1b, c. In neither analysis was *Diania* resolved as sister-taxon to Arthropoda, instead belonging to an unresolved polytomy at the base of the resultant most-parsimonious trees. To confirm that the trees of Liu *et al.*¹ are less parsimonious, a second pair of analyses was undertaken. Here tree topologies were constrained to the published strict-consensus topology (Fig. 1a), again using implied ($k = 2$) and equal weighting schemes. In both cases a single tree was found, substantially longer than those our first analyses recovered (16.4 as opposed to 12.4 steps under implied weighting; 119 as opposed to 89 steps under equal weighting). This demonstrates that the published topology is far from the most parsimonious that can be obtained from the data set. The use of a branch and bound (rather than heuristic) search indicates that this discrepancy is not a result of

chance failure to hit upon the best solution; instead we believe that it reflects a methodological error in the original analysis.

The phylogeny recovered here suggests a polyphyletic origin of arthropodized limbs, where the arthropodization in *Diania* was acquired separately to that of the crown-group arthropods. It does not rule out the ‘leg-first’ (as opposed to arthrodization-first) scenario posited by Liu *et al.*¹, but provides no support for it either. The recovered polychotomy demonstrates that the character coverage of Liu *et al.*¹ is inadequate to resolve the interrelationships of lobopodians; a more comprehensive data set is required to assess the position and significance of *Diania* properly. We note, however, that our results from the Liu *et al.*¹ data set (Fig. 1b) place the euarthropods within the dinocaridids (radiodonts and related taxa), and as such are similar to other analyses^{3,4} which have used more character- and taxon-rich data sets. The origin of arthropodized trunk limbs is problematical in this topology, as dinocaridids apparently lack any form of leg-like trunk appendage⁵. The increasingly detailed fossil record of stem-group euarthropods provides our best chance of resolving this issue, but as yet has failed to do so; unequivocal evidence for any particular ordering of acquisition in these characters is not yet available. *Diania* is a fascinating animal, but in its revised position it contributes little to this debate.

David A. Legg^{1,2}, Xiaoya Ma², Joanna M. Wolfe³,
Javier Ortega-Hernández⁴, Gregory D. Edgecombe² & Mark D. Sutton¹

¹Department of Earth Science and Engineering, Imperial College London, London SW7 2AZ UK.

e-mail: d.legg10@imperial.ac.uk

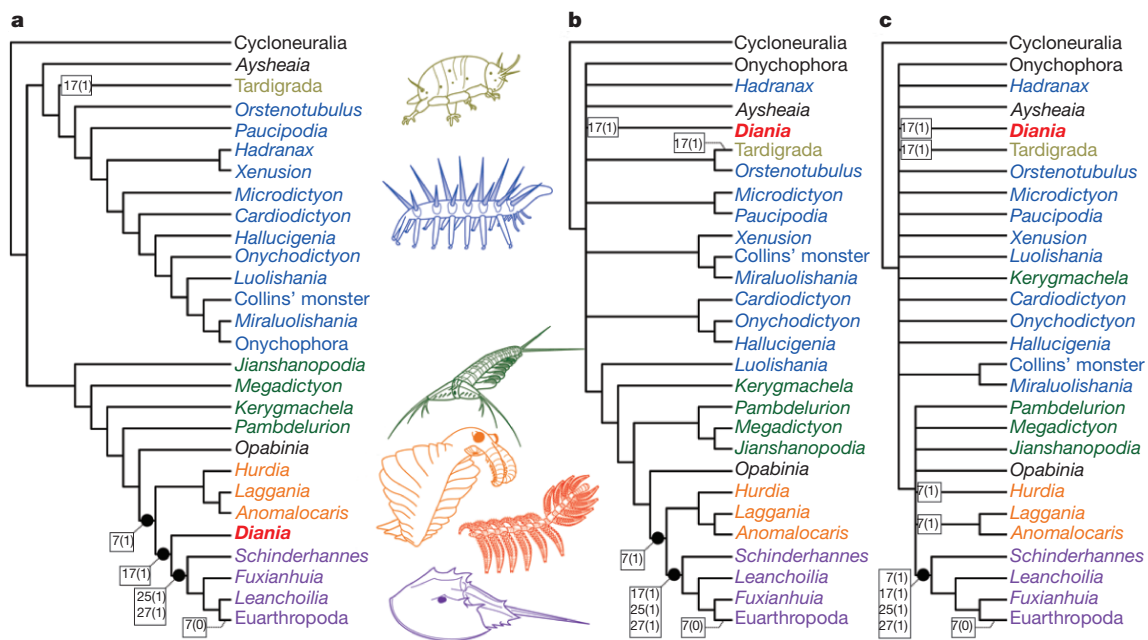


Figure 1 | Phylogeny of Cambrian lobopodians and stem-group arthropods. **a**, Topology proposed by Liu *et al.*¹ (implied weights, $k = 2$, 16.4 steps, consistency index (CI) = 0.336, retention index (RI) = 0.546; equal weights, 119 steps, CI = 0.336, RI = 0.546). **b**, Strict consensus of five most parsimonious trees using the methods of Liu *et al.*¹ (implied weights, $k = 2$, 12.4

steps, CI = 0.444, RI = 0.713). **c**, Strict consensus of 86 trees produced with equal character weighting (89 steps, CI = 0.499, RI = 0.718). Numbers represent important characters of Liu *et al.*¹ relating to arthropodization and arthrodization.

²Department of Palaeontology, Natural History Museum, London SW7 5BD, UK.

³Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA.

⁴Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK.

Received 3 March; accepted 24 May 2011.

1. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
2. Goloboff, P. *et al.* TNT, a free program for cladistic analysis. *Cladistics* **24**, 774–786 (2008).

3. Daley, A. C. *et al.* The Burgess Shale anomalocaridid *Hurdia* and its significance for early euarthropod evolution. *Science* **323**, 1597–1600 (2009).
4. Ma, X. *et al.* Morphology of *Luolishania longicruris* (Lower Cambrian, Chengjiang Lagerstätte, SW China) and the phylogenetic relationships within lobopodians. *Arth. Struct. Dev.* **38**, 271–291 (2009).
5. Zhang, X. & Briggs, D. E. G. The nature and significance of the appendages of *Opabinia* from the Middle Cambrian Burgess Shale. *Lethaia* **40**, 161–173 (2007).

Author Contributions: D.L. was responsible for writing and coordination of the manuscript; X.M., J.W., J.O.-H., G.E.D. and M.S. helped with discussion. D.L. and J.W. were also responsible for reanalysing the data. J.O.-H. produced the accompanying figure and M.S. aided with wording and stylistic editing of the original manuscript.

Competing financial interests: declared none.

doi:10.1038/nature10267

Liu *et al.* reply

REPLYING TO R. C. P. Mounce & M. Wills *Nature* **476**, doi:10.1038/nature10266 (2011); D. A. Legg *et al.* *Nature* **476**, doi:10.1038/nature10267 (2011);

We welcome the reanalyses by Mounce and Wills¹ and Legg *et al.*² of our paper³, and although we do not fully concur with their conclusions we are pleased that *Diania* has reopened the debate about key stages in arthropod evolution. We accept that the position of this fossil remains sensitive to parameters of analysis and in the original publication we conceded that our best-supported tree—*Diania* as sister-group to (*Schinderhannes* + Euarthropoda)—could be subject to

change, and that the ‘walking cactus’ may have a more basal position within the overall framework of the arthropod stem-group. These alternative treatments of our data would seem to confirm this suspicion, although we find the placement of *Diania* in an unresolved, and extremely basal, polytomy alongside velvet worms, tardigrades and various other lobopodians similarly problematical. We do not doubt that the authors’ results^{1,2} are statistically well supported, but what do these cladograms tell us about the evolution of the group? Lobopodians are, by their nature, fairly simple and consequently yield few convincing synapomorphies, either with each other or with arthropods in general. As we discovered, this makes scoring a robust data matrix including both lobopodians and arthropods challenging, and we wonder whether the basal polytomies recovered here are simply due to clustering among taxa with few unequivocal apomorphies and/or much missing data.

Our original placement of *Diania* close to the euarthropods was strongly influenced by the character of jointed trunk appendages. We acknowledge an error in scoring characters 17 and 27 of Tardigrada, which should have been 0 and 1, not 1 and 0, respectively. We also accept the criticism by Mounce and Wills¹ that parent–daughter characters are, to some extent, dependent upon one another and can artificially inflate support for particular clades. This situation is hard to avoid when selecting characters across a range of lobopod/arthropod fossils, which include taxa with unusual morphologies (for example, *Opabinia*) or where there are alternative hypotheses for the homology of a given lobopod feature and its probable arthropod equivalent: should a lobopod/dinocaridid flap be scored the same way as a euarthropod limb

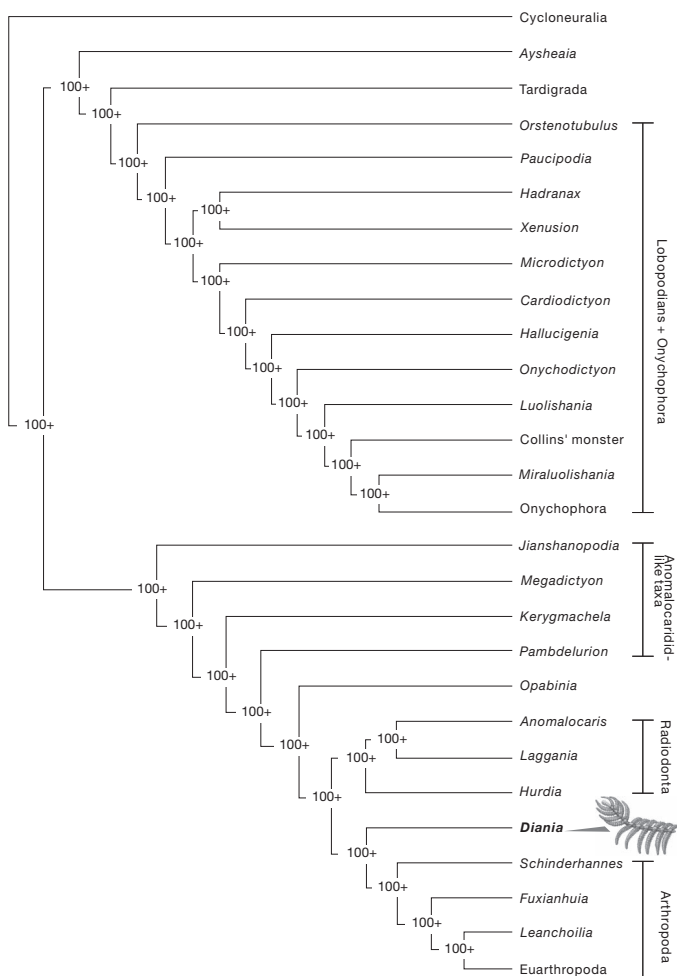


Figure 1 | Reanalysis of our data using PAUP: bootstrap 50% majority-rule consensus tree. Hierarchical structure in the data was assessed using the PTP test⁶ as implemented in PAUP. Parsimonious trees were found through a heuristic search strategy using tree bisection-reconnection (TBR) branch swapping. In an initial analysis, all characters were treated as unordered. Branch support was based on calculating bootstrap values (10,000 replicates and with rearrangement limit of 10,000,000 rearrangements per additional sequence). A significant value in the PTP test ($P = 0.01$) does suggest the presence of a phylogenetic signal in the morphological data, supporting the given topology. Thirty-eight characters were parsimony-informative. The heuristic search for a maximum parsimony solution resulted in only one parsimonious tree (length = 121, homoplasy index (HI) = 0.67, consistency index (CI) = 0.33, retention index (RI) = 0.55, excluding 1 uninformative character). This tree contains 26 resolved nodes, of which many were strongly supported by bootstrap values (100+) and in some nodes bootstrap values showed even greater resolution.

exite; could both be subsumed under a 'biramous limb', etc.? Furthermore, a disadvantage of PAUP is that 'gaps' are automatically treated as missing in the analysis, and the only way to distinguish between them is to treat the gap as a new state; although a gap in our analysis refers to an inapplicable character, which is itself not a new state.

In our initial analysis data were entered into a matrix using MacClade version 3.05 (ref. 4). Analyses were performed with PAUP version 4.0b10 (ref. 5), whereby multi-state characters were treated as 'unordered', other characters were treated as 'ordered', and a branch and bound search under implied weights ($k = 2$) produced many most parsimonious trees—including trees similar to those of Legg *et al.*² and Mounce and Wills¹. For the purpose of character state mapping, we selected one of these most parsimonious trees (see Fig. 4 in ref. 3) which we felt best reflected the complex relationships among lobopodians and arthropods. Also, a constrained analysis was conducted, and three monophyletic groups (lobopodians + Onychophora, Radiodonta, Arthropoda) were enforced. To verify the stability of this tree, we conducted bootstrap analysis and the result (Fig. 1) shows that the contained 26 resolved nodes of this tree were also strongly supported by bootstrap values (100+) and in some nodes bootstrap values

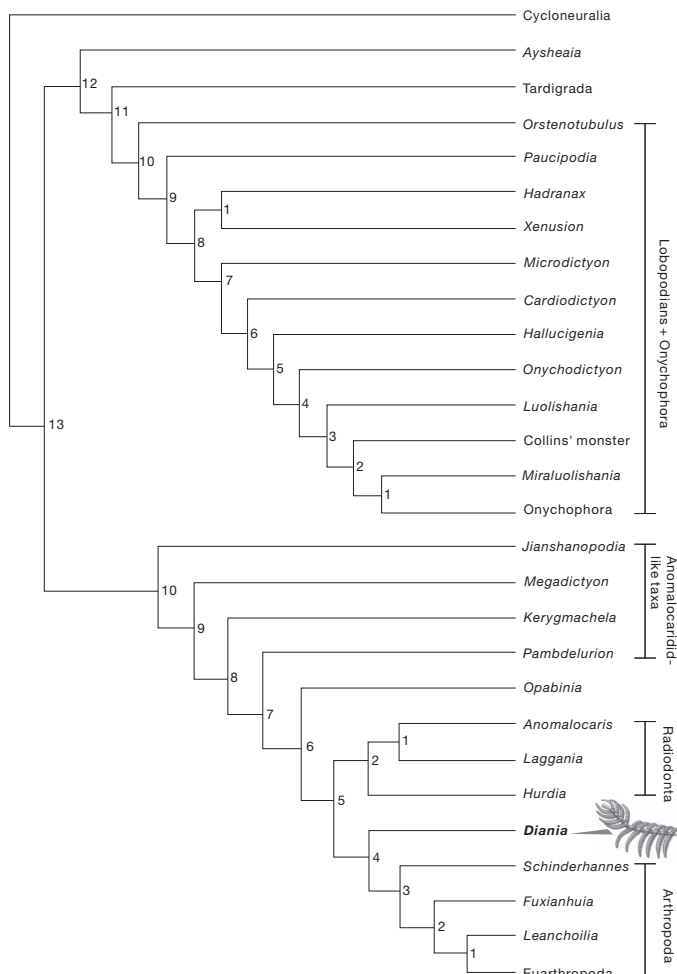


Figure 2 | Tree shows the Bremer support value, which strongly supports *Diania* being sister group to Arthropoda (Bremer index = 4). Four additional steps would be required to disrupt the *Diania*–arthropod relationship. As we noted in the original description, *Diania* is the most arthropod-like of the lobopodians recorded so far—at least in terms of limb morphology—and we feel that it is better placed on the arthropod stem, rather than in an unresolved polytomy.

showed more resolution. Additionally, a significant value of the partitioning tail permutation (PTP) test ($P = 0.01$) suggests the presence of a clear phylogenetic signal in the morphological data, also strongly supporting the topology shown. Furthermore, a test of Bremer support still supports our initial analysis (Fig. 2).

Legg *et al.*² drew attention to the absence of sclerotized trunk limbs in dinocaridids (*Anomalocaris*, etc.), which remains for us one of the great puzzles of stem-group arthropod evolution. Put simply, dinocaridids have a more arthropod-like head region (cephalisation, eyes, sclerotized mouthparts), whereas *Diania* lacks such sophistication in the anterior body region, but has jointed trunk appendages. One solution to this puzzle would be to assume that dinocaridids also had trunk limbs, but lost them secondarily. In this scenario *Diania* could sit comfortably as sister group to a (Dinocaridida + Arthropoda) clade: all three sharing jointed appendages, the latter two sharing cephalisation, etc.

We accept that the tree of Legg *et al.*² and Mounce and Wills¹ is the strict consensus tree, but we feel that the strict tree is, in this case, too conservative to provide meaningful information about the early evolution of the arthropods. This was undoubtedly a complex process, and may have involved numerous parallel developments introducing homoplasy. Under these circumstances, the comb-like most parsimonious tree is unhelpful and exemplifies a disadvantage of parsimony in this instance when faced with a complex early radiation. Indeed, Legg *et al.*² recognized that the strict consensus tree(s) “suggests a polyphyletic origin of arthropodized limbs”. If true, this would be significant. In this context, Mounce and Wills¹ seem to have overlooked the potential significance of their reanalysis of our data. Much evidence has been accumulated that Arthropoda is monophyletic, but the logical conclusion of both reanalyses is that jointed legs are homoplastic, at least within the lobopodian–arthropod assemblage. The implications of this are not trivial, bringing us close to the (largely discredited) Manton school of thought in which jointed appendages were proposed to have evolved in parallel in different (here euarthropod) lineages. At a fundamental level we need to know how easy it is to turn a soft, lobopodian limb into a sclerotized arthropod one, and whether this happened once, twice or on multiple occasions. For this reason alone, we believe that *Diania* is not merely fascinating, but remains invaluable to the evolutionary debate by challenging our notion of what it means to be an arthropod.

Jianni Liu^{1,2}, Michael Steiner², Jason A. Dunlop³, Helmut Keupp², Degan Shu^{1,4}, Qiang Ou⁴, Jian Han¹, Zhifei Zhang¹ & Xingliang Zhang¹

¹Early Life Institute, State Key Laboratory of Continental Dynamics, Department of Geology, Northwest University, Xi'an 710069, China.

²Department of Earth Science, Freie Universität Berlin, D-12249 Berlin, Germany.

e-mail: liu2009@zedat.fu-berlin.de or liujianni@126.com

³Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity at the Humboldt University Berlin, D-10115 Berlin, Germany.

⁴School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China.

1. Mounce, R. C. P. & Wills, M. Phylogeny of *Diania* challenged. *Nature* **476**, doi:10.1038/nature10266 (2011).
2. Legg, D. A. *et al.* Lobopodian phylogeny reanalysed. *Nature* **476**, doi:10.1038/nature10267 (2011).
3. Liu, J. *et al.* An armoured Cambrian lobopodian from China with arthropod-like appendages. *Nature* **470**, 526–530 (2011).
4. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution* Version 3 (Sinauer Associates, 1992).
5. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4.0. (Sinauer Associates, 2002).
6. Faith, D. P. & Cranston, P. S. Probability, parsimony, and Popper. *Syst. Biol.* **41**, 252–257 (1992).

doi:10.1038/nature10268

High plant diversity is needed to maintain ecosystem services

Forest Isbell¹, Vincent Calcagno¹, Andy Hector², John Connolly³, W. Stanley Harpole⁴, Peter B. Reich^{5,6}, Michael Scherer-Lorenzen⁷, Bernhard Schmid², David Tilman⁸, Jasper van Ruijven⁹, Alexandra Weigelt¹⁰, Brian J. Wilsey⁴, Erika S. Zavaleta¹¹ & Michel Loreau¹

Biodiversity is rapidly declining worldwide¹, and there is consensus that this can decrease ecosystem functioning and services^{2–7}. It remains unclear, though, whether few⁸ or many⁹ of the species in an ecosystem are needed to sustain the provisioning of ecosystem services. It has been hypothesized that most species would promote ecosystem services if many times, places, functions and environmental changes were considered⁹; however, no previous study has considered all of these factors together. Here we show that 84% of the 147 grassland plant species studied in 17 biodiversity experiments promoted ecosystem functioning at least once. Different species promoted ecosystem functioning during different years, at different places, for different functions and under different environmental change scenarios. Furthermore, the species needed to provide one function during multiple years were not the same as those needed to provide multiple functions within one year. Our results indicate that even more species will be needed to maintain ecosystem functioning and services than previously suggested by studies that have either (1) considered only the number of species needed to promote one function under one set of environmental conditions, or (2) separately considered the importance of biodiversity for providing ecosystem functioning across multiple years^{10–14}, places^{15,16}, functions^{14,17,18} or environmental change scenarios^{12,19–22}. Therefore, although species may appear functionally redundant when one function is considered under one set of environmental conditions⁷, many species are needed to maintain multiple functions at multiple times and places in a changing world.

Arguments for biodiversity conservation are often based on ecosystem services, but it remains unclear whether few⁸ or many⁹ species are needed to maintain ecosystem services. Determining how many species provide ecosystem services will require a synthesis of several areas of biodiversity research (Fig. 1). Biodiversity–ecosystem functioning studies have often considered a single functional context and found that multiple, but not all, study species promoted ecosystem functioning^{5–7} (Fig. 1a). We define a functional context (henceforth context) as the measurement of one function, at one time and place, under one environmental change scenario. Several related biodiversity studies have explored whether more species promote ecosystem functioning when more than one context is considered. For example, biodiversity–ecosystem stability (that is, the invariability of productivity) studies have found that more species are needed to provide ecosystem functioning at larger spatio-temporal scales because different species promote productivity at different times^{10–14} (Fig. 1b) or places^{15,16}. Biodiversity–ecosystem multifunctionality studies have found that more species are needed to provide multiple functions because different species promote different functions^{14,17,18} (Fig. 1c). Biodiversity–global change studies have found that more species are needed to

provide ecosystem functioning in a changing world because different species promote ecosystem functioning under different environmental change scenarios¹². Here, for the first time to our knowledge, we consider all of these relationships together.

We included data from 17 grassland biodiversity experiments that considered multiple times, places, functions or environmental change scenarios (Supplementary Table 1 and Supplementary Data). To test whether different species promoted ecosystem functioning during different years, we included studies that planted replicate plots (same species compositions) during consecutive years²³ or made repeated measurements of ecosystem functions across years^{17,19,21,22,24,25}. To test whether different species promoted ecosystem functioning at different places, we included studies that planted replicate plots (same species compositions, with one exception²⁵) at multiple sites across Europe²⁶ or multiple spatial blocks within a site^{17,23,25}. To test whether different

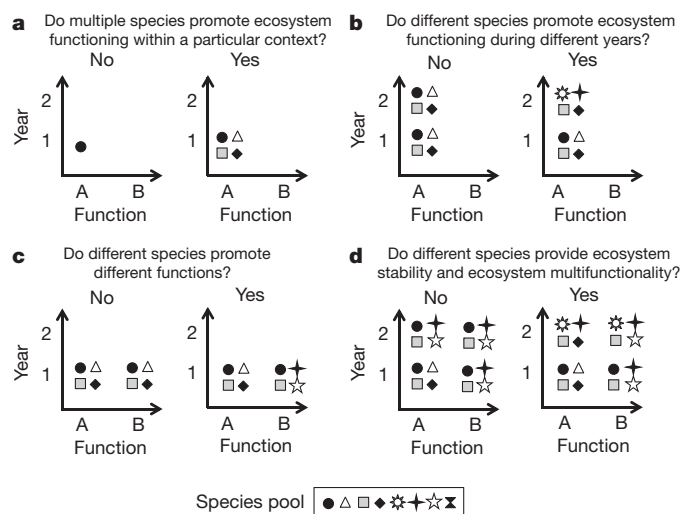


Figure 1 | Some of the ways that biodiversity can be important for ecosystem functioning. Each of the eight symbols represents a species. Species shown in the bivariate plots are those that promoted ecosystem functioning within each functional context (for example, context A1 might be above-ground biomass measured during 2001). Although this figure defines contexts in two dimensions for simplicity, we considered four dimensions (Fig. 3). Previous studies have considered (a) one context, or (b, c) one dimension of contexts (b, ecosystem stability studies; c, ecosystem multifunctionality studies). Figure 2 tests the question in d, by comparing the one-dimensional overlap (for example, between A1 and A2) with the two-dimensional overlap (for example, between A1 and B2) and three-dimensional overlap (that is, a pair of contexts that differs in three ways; not shown). Our results reject each of the null hypotheses shown on the left in a–d.

¹Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada. ²Institute of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057 Zurich, Switzerland. ³UCD School of Mathematical Sciences, University College Dublin, Dublin 4, Ireland. ⁴Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA. ⁵Department of Forest Resources, University of Minnesota, St Paul, Minnesota 55108, USA. ⁶Hawkesbury Institute for the Environment, University of Western Sydney, Richmond, New South Wales 2753, Australia. ⁷Faculty of Biology, University of Freiburg, Geobotany, D-79104 Freiburg, Germany. ⁸Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA. ⁹Nature Conservation and Plant Ecology, Wageningen University, 6700 AA Wageningen, The Netherlands. ¹⁰Institute of Biology I, University of Leipzig, 04103 Leipzig, Germany. ¹¹Environmental Studies Department, University of California, Santa Cruz, California 95064, USA.

species promoted different functions, we included studies that measured several functions^{14,17,22}, such as biomass production and nutrient uptake. Many of these functions are considered to be supporting ecosystem services because other types of ecosystem services depend on them^{4,27}. To test whether different species promoted ecosystem functioning under different environmental change scenarios, we included studies that applied environmental change treatments, such as nutrient and CO₂ enrichment¹⁹, precipitation changes²¹ or land use changes such as livestock grazing²² and haying²⁰.

We began by identifying the sets of study species that influenced ecosystem functioning within each context. Species were considered to promote ecosystem functioning in a particular context if they had effects in the direction that would usually be considered desirable from an ecosystem services perspective¹⁷. Positive effects were considered desirable for all functions except for soil inorganic nitrogen and light availability at ground level, where negative effects are consistent with lower levels of unconsumed resources¹⁷. We did not use separate definitions of desirable effects for different species (for example, positive effects of legumes on soil nitrogen might be considered desirable) to be consistent with previous studies^{17,18}, to be conservative and because it may not be possible to manage simultaneously for both positive and negative effects. We found that approximately 27% of the study species promoted ecosystem functioning within any particular context, regardless of the size of the study species pool (Fig. 2a). Note that if many species were functionally redundant, or if only the most common species promoted ecosystem functioning, then we would expect a saturating relationship in Fig. 2a. Instead, our results suggest that even rare species can promote ecosystem functioning.

After identifying the sets of species that promoted ecosystem functioning in each context, we tested whether different sets of species promoted ecosystem functioning in different contexts. We used Sørensen's similarity index to quantify overlap between species sets¹⁷. All comparisons were made within studies so that differences between pairs of contexts were not due to sampling from multiple species pools. First, we quantified one-dimensional overlap between all pairs of contexts that differed in only one way (Fig. 1b, c). For example, multi-year overlap was quantified between each pair of contexts that differed only in which year ecosystem functioning was measured (that is, same

place, function and environmental change scenario in both contexts). A multi-year overlap value of one or zero would respectively indicate that completely identical or completely unique sets of species promoted ecosystem functioning during different years, independent of the other sources of variation. We found overlap values between these two extremes, which indicates that somewhat different sets of species promoted ecosystem functioning during different years, at different places, for different functions and under different environmental change scenarios (Fig. 2b).

After considering these sources of variation independently, we quantified multi-dimensional overlap between pairs of contexts that differed in two or three ways (Fig. 1d). Again, all comparisons were made within studies. We found that the average overlap between pairs of contexts decreased as the number of differences between contexts increased (Fig. 2b and Supplementary Fig. 1). This means that, for example, the identities of the additional species needed to provide one function during multiple years were not the same as the identities of the additional species needed to provide multiple functions during one year (Fig. 1d). Additionally, species sets did not simply vary independently of context attributes (permutation test $P < 0.001$ for one-, two- and three-dimensional overlap) (Fig. 2b). Thus, our results indicate that even more species will be needed to maintain ecosystem functioning and services than previously suggested by studies that have either (1) considered only the number of species needed to promote one function under one set of environmental conditions, or (2) separately considered the importance of biodiversity for providing ecosystem functioning across multiple years^{10–14}, places^{15,16}, functions^{14,17,18} or environmental change scenarios^{12,19–22}. Future studies could more completely consider the consequences of biodiversity declines for ecosystem functioning and services by similarly considering the multi-dimensionality of ecosystem functioning both in experimental and natural communities.

Next, we quantified the extent to which the number of species promoting ecosystem functioning increased as more years, places, functions or environmental change scenarios were considered within each study. In other words, we quantified the accumulation of species across each of the four dimensions of contexts that we considered (Fig. 1). We found that a greater proportion of species promoted

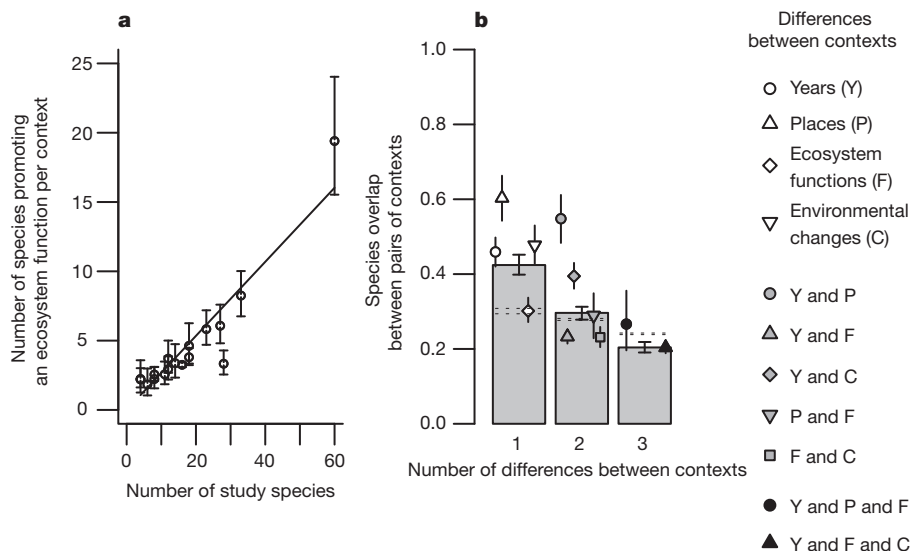


Figure 2 | Sets of study species that promoted ecosystem functioning. **a**, The mean number of species that promoted ecosystem functioning within each context increased linearly ($t = 16.40$, $P < 0.001$, $R^2 = 0.944$) with the size of the species pool, such that approximately 27% (mean, 95% confidence intervals for slope: 0.27, 0.24–0.30) of the study species promoted ecosystem functioning within each context. Error bars for each study indicate 95% generalized linear model confidence intervals. **b**, Different sets of species promoted ecosystem

functioning in different contexts (overlap < 1), and overlap between pairs of contexts decreased as the number of differences between contexts increased (see Fig. 1). Symbols indicate means for each specific type of overlap; horizontal dotted lines show $\pm 95\%$ permutation test confidence intervals; error bars for symbols and bars indicate 95% bootstrap confidence intervals. Supplementary Data indicates numbers of contexts for each study.

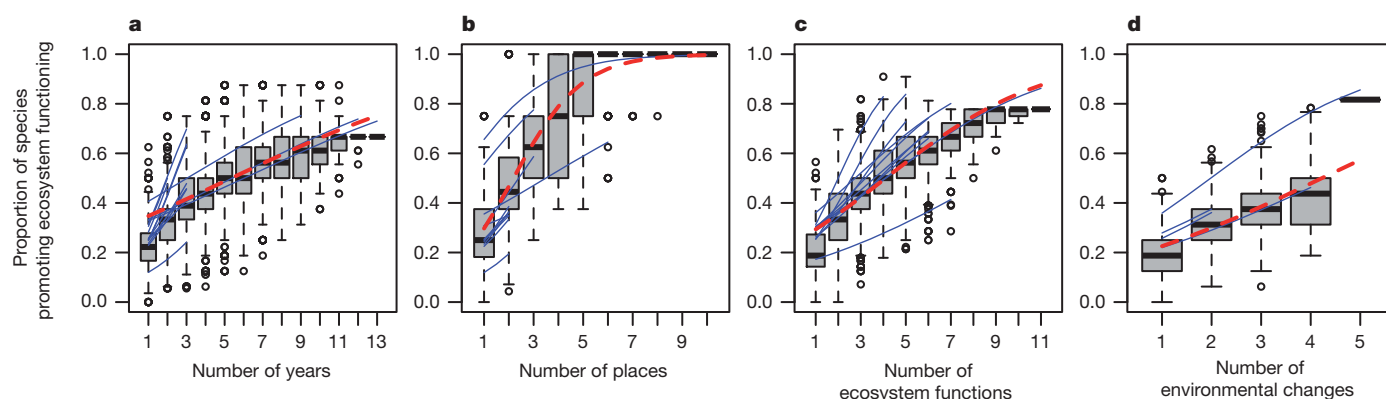


Figure 3 | The proportion of study species that promoted ecosystem functioning increased when more (a) years, (b) places, (c) ecosystem functions and (d) environmental change scenarios were independently considered. Solid blue lines indicate generalized linear model fits for each study; dashed red line indicates grand mean generalized linear model fitted

ecosystem functioning when more years, places, functions or environmental change scenarios were considered (Fig. 3). These relationships result from different species promoting ecosystem functioning in different contexts (Fig. 2b). Note that if the one-dimensional overlap values corresponding to each panel in Fig. 3 were one or zero, then these relationships would be horizontal or linearly increasing, respectively¹⁷. Our results are between these two extremes.

After comparing contexts within studies, data from all studies were combined to consider how the total number of species that promoted ecosystem functioning increased with the total number of contexts. We quantified the number of species that promoted ecosystem functioning in a random subset of all possible combinations of our observed contexts (that is, 100 pairs, 100 groups of three, etc.). The large increase in the number of species that promoted ecosystem functioning as more contexts were considered (Fig. 4) is the result of different species promoting ecosystem functioning during different years, at different places, for different functions, under different environmental change scenarios and in different species pools. Considering all of these factors together suggests that many species will be needed to maintain ecosystem multifunctionality at large spatio-temporal scales in a changing world. Consequently, the extinction (or decreased local occurrence) of almost any of these species is expected to decrease ecosystem functioning and services in at least some contexts. Further study is needed to identify the processes that explain why different species promoted ecosystem functioning in different contexts. The specific mechanisms involved probably differed across contexts, but previous results from these and other biodiversity experiments²⁸ suggest that complementarity (in time, space, functional effect traits and functional response traits) is a general explanation for this pattern.

Our results reveal new opportunities and challenges for prioritizing conservation efforts and predicting consequences of biodiversity declines. According to the precautionary principle, all species should be conserved because we cannot be certain which species actually provide ecosystem services²⁹. Our results offer further support for the precautionary principle because most of the studied species were important at least once, and species exhibited context-dependent effects that will be difficult to predict. If it is impossible or impractical to conserve all species, then future studies could additionally consider how often (Supplementary Fig. 2 and Supplementary Data) and how much species influenced ecosystem functioning to determine which species are most important for maintaining ecosystem functioning and services. This will require careful consideration of many contexts, because it is not possible to make general predictions or conclusions by considering a few context-dependent phenomena³⁰.

Future studies could determine whether some species consistently promote ecosystem functioning under environmental conditions that

are currently common, or under environmental change scenarios that will probably become increasingly common. Note that even species that have small effects could be important for maintaining ecosystem functioning and services if they have a large cumulative desirable effect across many contexts. For example, *Eriochloa sericea* had the smallest desirable effect on above-ground biomass in the irrigated plots at the MEND Irrigation experiment during 2009, but promoted ecosystem functioning in 75% of the contexts in which it was included. Future studies could also determine which species promote ecosystem functioning in particular contexts that are highly valued by stakeholders. Note that even species that rarely promote (or often decrease) ecosystem functioning could be most important for maintaining ecosystem functioning and services within some contexts. For example, although *Pascopyrum smithii* only promoted ecosystem functioning in 2% of the contexts in which it was included, it promoted soil carbon more than any other species in the Cedar Creek Biodiversity experiment during 2004. Furthermore, note that declines in local diversity, which are far more common than global extinctions, will also decrease ecosystem functioning and services within some contexts. Finally, even the few species that never promoted ecosystem functioning in these studies (Supplementary Fig. 2) could promote ecosystem functioning

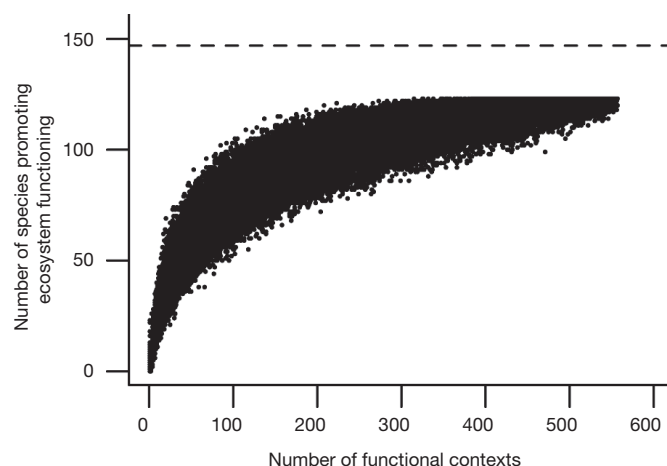


Figure 4 | The number of study species that promoted ecosystem functioning increased with the number of contexts considered across all studies. The points are the number of species that promoted ecosystem functioning when 1–557 contexts were sampled from all 557 contexts. The dashed line indicates the total number of studied species (147), which restricts the upper limit for these values. The x axis includes variation across years, places, functions, environmental change scenarios and species pools.

in other contexts, or be a conservation priority for other (for example, ethical, aesthetic) reasons. Therefore, we encourage careful consideration of many contexts when making conservation decisions and predicting the consequences of biodiversity declines.

METHODS SUMMARY

Identifying the species that promoted ecosystem functioning. For each context, we modelled ecosystem functioning response variables as a function of the presence or absence of each study species at the plot level. We used backward-elimination multiple regression to identify a minimally adequate model, based on the Akaike information criterion (AIC)¹⁷. This procedure was performed with the stepAIC function in the MASS package of R 2.11.1 (see Methods). Species were considered to promote ecosystem functioning in a context if they were included in the minimally adequate model and had effects in the direction that would usually be considered desirable from an ecosystem services perspective¹⁷.

Comparing sets of species between pairs of contexts. We used Sørensen's similarity index to quantify the overlap between the sets of species that promoted ecosystem functioning in pairs of contexts within each study¹⁷. This allowed us to test whether identical (overlap = 1), unique (overlap = 0) or somewhat different ($0 < \text{overlap} < 1$) sets of species promoted ecosystem functioning in different contexts. We also tested whether overlap decreased as the number of differences between contexts increased (Fig. 1).

Accumulation of species across contexts. We quantified the accumulation of species across each of the four dimensions of contexts that we considered (Methods; Fig. 1). A quasi-binomial generalized linear model was fitted to determine how the proportion of species that promoted ecosystem functioning increased with the number of years (or places, functions, environmental changes), including 'study' as a factor. This allowed us to describe the mean trends within and across studies. We also randomly sampled combinations (that is, 100 pairs, 100 groups of three, and so on) of all 557 contexts to determine how the number of species promoting ecosystem functioning increased when all of these factors were considered together.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 March; accepted 10 June 2011.

Published online 10 August 2011.

- Butchart, S. H. M. *et al.* Global biodiversity: indicators of recent declines. *Science* **328**, 1164–1168 (2010).
- Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808 (2001).
- Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
- Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
- Cardinale, B. J. *et al.* Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* **443**, 989–992 (2006).
- Naeem, S., Bunker, D. E., Hector, A., Loreau, M. & Perrings, C. *Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective* (Oxford Univ. Press, 2009).
- Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572–592 (2011).
- Ridder, B. Questioning the ecosystem services argument for biodiversity conservation. *Biodivers. Conserv.* **17**, 781–790 (2008).
- Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ.* **7**, 437–444 (2009).
- McNaughton, S. J. Diversity and stability of ecological communities: a comment on the role of empiricism in ecology. *Am. Nat.* **111**, 515–525 (1977).
- Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
- Craine, J. M. *et al.* The role of plant species in biomass production and response to elevated CO₂ and N. *Ecol. Lett.* **6**, 623–630 (2003).
- Tilman, D., Reich, P. B. & Knops, J. M. H. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441**, 629–632 (2006).

- Zavaleta, E. S., Pasari, J. R., Hulvey, K. B. & Tilman, G. D. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proc. Natl Acad. Sci. USA* **107**, 1443–1446 (2010).
- Loreau, M., Mouquet, N. & Gonzalez, A. Biodiversity as spatial insurance in heterogeneous landscapes. *Proc. Natl Acad. Sci. USA* **100**, 12765–12770 (2003).
- Griffin, J. N. *et al.* Spatial heterogeneity increases the importance of species richness for an ecosystem process. *Oikos* **118**, 1335–1342 (2009).
- Hector, A. & Bagchi, R. Biodiversity and ecosystem multifunctionality. *Nature* **448**, 188–190 (2007).
- Garnfeldt, L., Hillebrand, H. & Jonsson, P. R. Multiple functions increase the importance of biodiversity for overall ecosystem functioning. *Ecology* **89**, 1223–1231 (2008).
- Reich, P. B. *et al.* Plant diversity enhances ecosystem responses to elevated CO₂ and nitrogen deposition. *Nature* **410**, 809–812 (2001).
- Weigelt, A., Weisser, W. W., Buchmann, N. & Scherer-Lorenzen, M. Biodiversity for multifunctional grasslands: equal productivity in high-diversity low-input and low-diversity high-input systems. *Biogeosciences* **6**, 1695–1706 (2009).
- Wilsey, B. J., Teaschner, T. B., Daneshgar, P. P., Isbell, F. I. & Polley, H. W. Biodiversity maintenance mechanisms differ between native and novel exotic-dominated communities. *Ecol. Lett.* **12**, 432–442 (2009).
- Isbell, F. I. & Wilsey, B. J. Increasing native, but not exotic, biodiversity increases aboveground productivity in ungrazed and intensely grazed grasslands. *Oecologia* **165**, 771–781 (2011).
- Isbell, F. I., Losure, D. A., Yurkonis, K. A. & Wilsey, B. J. Diversity–productivity relationships in two ecologically realistic rarity and extinction scenarios. *Oikos* **117**, 996–1005 (2008).
- Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
- van Ruijven, J. & Berendse, F. Long-term persistence of a positive plant diversity–productivity relationship in the absence of legumes. *Oikos* **118**, 101–106 (2009).
- Kirwan, L. *et al.* Evenness drives consistent diversity effects in intensive grassland systems across 28 European sites. *J. Ecol.* **95**, 530–539 (2007).
- Millennium Ecosystem Assessment. *Ecosystems and Human Well-being: Synthesis* (Island Press, 2005).
- Cardinale, B. J. *et al.* Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proc. Natl Acad. Sci. USA* **104**, 18123–18128 (2007).
- Ehrlich, P. & Ehrlich, A. *Extinction: The Causes and Consequences of the Disappearance of Species* (Victor Gollancz, 1982).
- Lawton, J. H. Are there general laws in ecology? *Oikos* **84**, 177–192 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Byrnes, L. Gamfeldt and M. Emmerson for comments on an earlier version of this manuscript. We thank the Swiss SystemsX.ch initiative (IPP-2008/23) for supporting this project. The BIODEPTH project was funded by the European Commission within the Framework IV Environment and Climate Programme (ENV-CT95-0008) and by the Swiss Federal Office for Education and Science (Project EU-1311). The Jena Experiment was funded by the Deutsche Forschungsgemeinschaft (DFG, FOR 456), Friedrich Schiller University of Jena, Max Planck Society, University of Zurich, Swiss National Science Foundation (3100AO-107531) and ETH Zurich. The Wageningen experiment was funded by the Dutch Organisation for Scientific Research (NWO) within the framework of the Biodiversity Programme. Work on the Agrobiodiversity experiment was funded by the EU Commission through COST Action 852 and Science Foundation Ireland (09/RFP/E0B2546). The BioCON experiment was funded by the US Department of Energy (DOE/DE-FG02-96ER62291) and the US National Science Foundation (Biocomplexity 0322057, LTER DEB 9411972, DEB 0080382, DEB 0620652 and LTREB DEB 0716587). The MEND Irrigation, BioGEN and Rarity–Extinction experiments were funded by the US National Science Foundation (DEB 0639417). The Cedar Creek Biodiversity experiment was funded by the US National Science Foundation. M.L. was supported by The Natural Sciences and Engineering Research Council of Canada (Discovery Grant) and the Canada Research Chair program.

Author Contributions F.I. conceived the project; J.C., A.H., F.I., P.B.R., M.S.-L., B.S., D.T., J.v.R., A.W. and B.J.W. designed and conducted experiments; F.I. and V.C. analysed the data, with input from A.H. and M.L.; F.I. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.I. (forest.isbell@gmail.com).

METHODS

Identifying the species that promoted ecosystem functioning. For each context, we modelled ecosystem functioning response variables as a function of the presence or absence of each study species at the plot level. We used a backward-elimination multiple regression analysis to identify a minimally adequate model, based on the AIC¹⁷. This procedure was performed with the stepAIC function in the MASS package of R 2.11.1. Specifically, for each context, we started with a full model that included a main effect for each study species and an intercept. The stepAIC function then removed each species, one at a time, from this full model and compared the AIC values of the resulting simpler models with the AIC value of the full model. If the AIC value for any of the simpler models was smaller than the AIC value for the full model, then the variable whose removal resulted in the largest decrease in AIC was permanently removed from the full model. This backward-deletion process was repeated until the removal of any species resulted in a model with a higher AIC value. The minimally adequate model that resulted from this process contained the most parsimonious set of species influencing ecosystem functioning. These species were considered to promote ecosystem functioning if they had effects in the direction that would usually be considered desirable from an ecosystem services perspective¹⁷. Positive effects were considered desirable for all functions except for soil inorganic nitrogen and light availability at ground level, where negative effects are consistent with lower levels of unconsumed resources¹⁷. We did not use separate definitions of desirable effects for different species (for example, positive effects of legumes on soil nitrogen might be considered desirable) to be consistent with previous studies^{17,18}, to be conservative and because it may not be possible to manage simultaneously for both positive and negative effects.

This modelling approach is conservative in several ways. Previous approaches for determining the number of species that promoted ecosystem functioning could have been biased by not allowing species to decrease ecosystem functioning¹⁸ or by not identifying which species actually influenced ecosystem functioning¹⁴. Furthermore, including species interactions in these or other models would probably increase the proportion of study species that promoted ecosystem functioning within each context^{12,26}. For example, some species perform poorly in monocultures, but interact positively with other species in mixtures (for example, grass–legume interactions)³¹. Our approach would underestimate the desirable effects of these species. Also, the presence or absence model that we used would be especially conservative for experimental designs that concentrate on varying evenness rather than richness, such as the simplex design used in the Agrobiodiversity study²⁶. Additionally, our results were qualitatively similar, but less conservative, when we used a model-averaging approach implemented in the glmulti package (version 0.6-3) of R³².

To compare the number of species that promoted ecosystem functioning across studies, we used a quasi-Poisson generalized linear model with ‘study’ as a main effect. This procedure was performed with the generalized linear model function in the stats package of R. The quasi-maximum-likelihood version of the Poisson generalized linear model accounts for over- or under-dispersion in the data. We used the confint.glm function in the MASS package of R to obtain the 95% confidence intervals. To quantify the proportion of study species that promoted ecosystem functioning, we regressed the mean number of species that promoted ecosystem functioning per context on the number of study species (that is, the number of planted species that were present at least once in biomass samples) with no intercept.

Comparing sets of species between pairs of contexts. After identifying the sets of species that promoted ecosystem functioning in each context, we quantified the overlap between these sets of species to test whether different sets of species promoted ecosystem functioning in different contexts. All overlap comparisons were made within studies. Overlap between contexts *a* and *b* was quantified by Sørensen’s similarity index¹⁷:

$$o = \frac{|E_a \cap E_b|}{0.5(|E_a| + |E_b|)}$$

where $|E_a|$ is the number of species that promoted ecosystem functioning in context *a* and $|E_a \cap E_b|$ is the number of species that promoted ecosystem functioning in both contexts. First, one-dimensional overlap was quantified between pairs of contexts that only differed in one way (Fig. 1b, c). For example, multi-year

overlap was quantified between each pair of contexts that differed only in which year ecosystem functioning was measured (that is, place, function and environmental change scenarios were the same in both contexts). This allowed us to test whether identical (overlap = 1), unique (overlap = 0) or somewhat different ($0 < \text{overlap} < 1$) sets of species promoted ecosystem functioning in different contexts. Next, multi-dimensional overlap was quantified between pairs of contexts that differed in two or three ways (Fig. 1d). This allowed us to test whether overlap decreased as the number of differences between contexts increased (Fig. 1).

For each type of overlap, we used non-parametric bootstrap, with correction for bias^{33,34}, to build 95% confidence intervals. We re-sampled the observed contexts with replacement to generate each of 1,000 bootstrap data sets. For each of these bootstrap data sets, we computed the average overlap across all pairs of contexts that were relevant (that is, for the particular type of overlap being considered). We used this bootstrap approach to control for the non-independence of pairwise comparisons that had one context in common. We then used an exact permutation approach³⁴ to test whether the observed overlap values differed from the expected null value (the null hypothesis being that overlap varied across contexts independent of context attributes: year, place, function and environmental change scenario). Within each study, we permuted context data (that is, the sequence of ones or zeros indicating whether each species had a desirable effect or not) with respect to context attributes. For each of 1,000 permutations, we computed the average overlap across all pairs of contexts that were relevant (that is, for the particular type of overlap being considered). We used the null distribution of overlap obtained from these random permutations to test whether the observed value was significantly lower (that is, in the lowest 2.5 percentiles) or significantly higher (that is, in the top 2.5 percentiles) than the null. We used this test because it does not assume that study species had independent responses to context attributes, and thus it should be robust to the presence of functional groups.

Note that there could be correlations between contexts owing to repeated measurements across years. These correlations could influence overlap estimates and tests that assume independent observations. For example, positive correlations between repeated measurements during consecutive years could lead to (1) over-estimates of multi-year (Y) overlap, and (2) smaller estimates of overlap between two different years at two different places than between two different years at one place (that is, Y and P overlap less than Y overlap). These potential correlations had little influence on our results because multi-year overlap estimates were much less than 1, and Y and P overlap estimates were not less than Y overlap estimates (Fig. 2). Nevertheless, we encourage future studies to consider these correlations when interpreting overlap estimates.

Accumulation of species across contexts. We quantified the accumulation of species across each of the four dimensions of contexts that we considered (Fig. 1). For example, above-ground biomass was sampled during 13 years at the Cedar Creek Biodiversity experiment. To determine the extent to which more species promoted above-ground biomass as more years were considered, we sampled all combinations of these 13 contexts (that is, all pairs, groups of three, etc.), and recorded the number of unique species that promoted ecosystem functioning, and the number of unique years, for each combination. This was repeated for each function, at each place, under each environmental change scenario. These results are summarized in Fig. 3a. A quasi-binomial generalized linear model was fitted to determine how the proportion of species that promoted ecosystem functioning increased with the number of years, including ‘study’ as a factor. This allowed us to describe the mean trends within and across studies. We also randomly sampled combinations (that is, 100 pairs, 100 groups of three, etc.) of all 557 contexts to determine how the number of species promoting ecosystem functioning increased when all of these factors were considered together.

1. Nyfeler, D. *et al.* Strong mixture effects among four species in fertilized agricultural grassland led to persistent and consistent transgressive overyielding. *J. Appl. Ecol.* **46**, 683–691 (2009).
2. Calcagno, V. & de Mazancourt, C. glmulti: an R package for easy automated model selection with (generalized) linear models. *J. Stat. Softw.* **34**, 1–29 (2010).
3. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–77 (1986).
4. Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* 2nd edn (Chapman and Hall, 1997).

Polar actomyosin contractility destabilizes the position of the cytokinetic furrow

Jakub Sedzinski^{1,2*}, Maté Biro^{1,2*}, Annelie Oswald¹, Jean-Yves Tinevez^{1,2†}, Guillaume Salbreux³ & Ewa Paluch^{1,2}

Cytokinesis, the physical separation of daughter cells at the end of mitosis, requires precise regulation of the mechanical properties of the cell periphery^{1,2}. Although studies of cytokinetic mechanics mostly focus on the equatorial constriction ring³, a contractile actomyosin cortex is also present at the poles of dividing cells^{2,4}. Whether polar forces influence cytokinetic cell shape and furrow positioning remains an open question. Here we demonstrate that the polar cortex makes cytokinesis inherently unstable. We show that limited asymmetric polar contractions occur during cytokinesis, and that perturbing the polar cortex leads to cell shape oscillations, resulting in furrow displacement and aneuploidy. A theoretical model based on a competition between cortex turnover and contraction dynamics accurately accounts for the oscillations. We further propose that membrane blebs, which commonly form at the poles of dividing cells⁵ and whose role in cytokinesis has long been enigmatic, stabilize cell shape by acting as valves releasing cortical contractility. Our findings reveal an inherent instability in the shape of the dividing cell and unveil a novel, spindle-independent mechanism ensuring the stability of cleavage furrow positioning.

Cell division is driven by the constriction of an actomyosin ring at the cell equator. However, throughout cytokinesis, a substantial actomyosin cortex remains in regions outside of the equatorial ring (the cell poles), both in cultured cells⁶, and *in vivo*^{7,8} (Supplementary Fig. 1 and Supplementary Movie 1). Mechanical studies indicate that the polar cortex regulates cleavage dynamics by generating resistive forces that slow furrow ingression⁹ (Supplementary Model section 1). The polar cortex is also able to exert stresses, for example to correct for shape perturbations¹⁰. Moreover, the formation of membrane blebs, known to be caused by cortical contractions, is commonly observed at the poles of dividing cells⁵ (Supplementary Movie 1). However, how polar contractile forces contribute to the overall cytokinetic cell shape is poorly understood.

During asymmetric division, unbalanced contraction of one of the poles has been shown to directly generate different-sized daughters in *Caenorhabditis elegans* QR.a neuroblasts¹¹. In symmetrically dividing cells, depletion of the actin-binding proteins anillin^{12–14}, supervillin¹⁵, profilin or diaphanous¹⁶, as well as astral microtubule stabilization¹⁷, result in improper polar contractions, leading to oscillations of the furrow position and division failure. It is, however, not known whether the poles influence furrow positioning in untreated cells. Reasoning that if substantial contractile forces are exerted at the poles, shape instabilities might arise (Fig. 1a), we asked whether a precise balance of polar forces is necessary for accurate symmetric division and, if so, how the cell achieves such a balance.

We first investigated whether asymmetries in polar forces could be observed during cytokinesis. To this aim, we quantified cytoplasm movements in dividing HeLa and L929 fibroblasts, as an imbalance in polar contractions would result in cytoplasm transfer between the poles. We detected cytoplasmic flows of amplitude at least two times higher than noise level (estimated using metaphase cells as a reference)

in 20% of HeLa ($n = 20$) and 25% of L929 cells ($n = 40$) (Fig. 1b, c, Supplementary Fig. 2 and Supplementary Movie 2). In most cases, the direction of the flows alternated, resulting in small oscillations. Strikingly, observations of a large number of dividing cells revealed that in about 2% of HeLa ($n = 282$) and L929 ($n = 526$) cells (Supplementary Table 1), dramatic shape oscillations occurred, where each pole fully contracted before reversal of the flow direction, destabilizing the equatorial position of the furrow and leading to division failure (Supplementary Movie 3). To test whether the shape instabilities were driven by cortex contractions, we used Y27632, a Rho-kinase inhibitor¹⁸, that, at low doses, decreases myosin activity, and thus cortical tension¹⁸, without preventing cytokinesis¹⁹. No cytoplasmic flows could be detected during cytokinesis in Y27632-treated cells (Fig. 1b, c and Supplementary Fig. 2, $n = 10$), supporting the view that cortical contractions drive the observed flows. We then tested the involvement of spindle microtubules, which influence polar contractility^{20,21}. Spindle disassembly with nocodazole increased the occurrence of shape instabilities (Supplementary Fig. 3, Supplementary Movie 4 and Supplementary Table 1), suggesting that the spindle is unlikely to drive cytokinetic oscillations. Taken together, our observations suggest that actomyosin contractions cause shape instabilities during cytokinesis, which can lead to division failure if their amplitude is not limited.

To assess which factors may trigger shape instabilities, we developed a minimal physical model of a cytokinetic cell. The dividing cell with a cortex under contractile tension is reminiscent of two connected soap bubbles, where Laplace pressure resulting from surface tension is known to lead to instabilities²² (Fig. 1d). Similarly, in a dividing cell, cortical tension generates hydrostatic pressure at each pole⁹. However, the cell interior is a poro-elastic material²³ that resists deformations and could stabilize cell shape against cortex-driven instabilities (Fig. 1d). In a first approximation, we assumed that each pole has a constant tension T and a linear bulk elasticity K (Supplementary Fig. 4a and Supplementary Model section 2), which accounts for the effects of cellular elastic components and of substrate adhesions (Supplementary Fig. 5, Supplementary Movie 5 and Supplementary Model section 5). Previous mechanical models of the polar cortex in cytokinesis have focused on symmetric cells^{6,9}, while we allowed variations in pole sizes. A stability analysis reveals that above a critical value of a stability parameter T/R_0K (with R_0 the radius of one daughter cell), any fluctuation in volume or tension destabilizes the symmetric cell shape and leads to cytosol flows between the poles (Supplementary Fig. 4b). Moreover, if a large enough difference in tension is introduced between the two poles, no bipolar solution for the shape of the cell can be found (Supplementary Fig. 4c).

To test whether an imbalance in contractile forces between the poles indeed destabilizes cell shape, we artificially introduced such a tension imbalance. Using a double-pipette local delivery setup (Fig. 2a), we disrupted the cortex at one pole of dividing HeLa cells by short application of the actin-depolymerizing drug cytochalasin D. Local polar cortex disruption resulted in the formation of a large bleb and destabilized the

¹Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. ²International Institute of Molecular and Cell Biology, 02109 Warsaw, Poland. ³Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany. [†]Present address: Pasteur Institute, 75724 Paris, France.

*These authors contributed equally to this work.

shape of the cell, triggering extensive shape oscillations (Fig. 2b and Supplementary Movie 6; see Supplementary Table 1). As an alternative method for cortex disruption, we used laser ablation¹⁸ (Fig. 2c). Polar cortex ablation also triggered bleb formation, followed by shape oscillations and division failure in a large portion of cytokinetic HeLa (Fig. 2d and Supplementary Movie 7, 35% of 20 cells) and L929 cells (50% of 20 cells). These experiments indicate that introducing a tension imbalance between the poles of a cleaving cell can trigger shape instabilities.

Our physical description also predicts that a global polar tension increase favours shape instabilities (Supplementary Fig. 4b). Cytokinetic shape oscillations had previously been observed upon depletion of the actin-binding scaffold protein anillin^{12–14}. Anillin depletion affects the

cortex globally; it is thus unlikely to lead to a strong inhomogeneity in tension between the poles. We analysed cortex distribution along the cell contour in HeLa and L929 cells expressing myosin regulatory light chain (MRLC)-tandem-dimer red fluorescent protein (tdRFP) and Lifeact-green fluorescent protein (GFP) to mark F-actin²⁴. We found that in anillin-depleted cells, the cortex was less enriched at the equator than in control cells, suggesting an enhanced polar contractility before shape oscillations (Fig. 2e–g, Supplementary Fig. 6 and Supplementary Movie 8). Another example of global cortex reinforcement was observed after cortex ablation in L929 cells, where oscillations could be triggered by ablation in any region of the cortex (Supplementary Fig. 7a and Supplementary Movie 9, 55% of 60 cells). Laser ablation in these cells led to cycles of cortex disassembly and reassembly, and resulted in an enhancement of the polar cortex at the expense of the equator (Supplementary Fig. 7). Taken together, these observations support the hypothesis that a global polar cortex reinforcement favours shape instabilities.

Our initial theoretical description predicted that the dividing cell shape should be unstable; however, it did not predict shape oscillations, which we systematically observe upon polar cortex perturbation. To gain insight into the oscillation mechanism, we analysed cortex and pole size dynamics during oscillations in anillin-depleted and laser-ablated cells. The mean fluorescence intensities of cortical Lifeact and

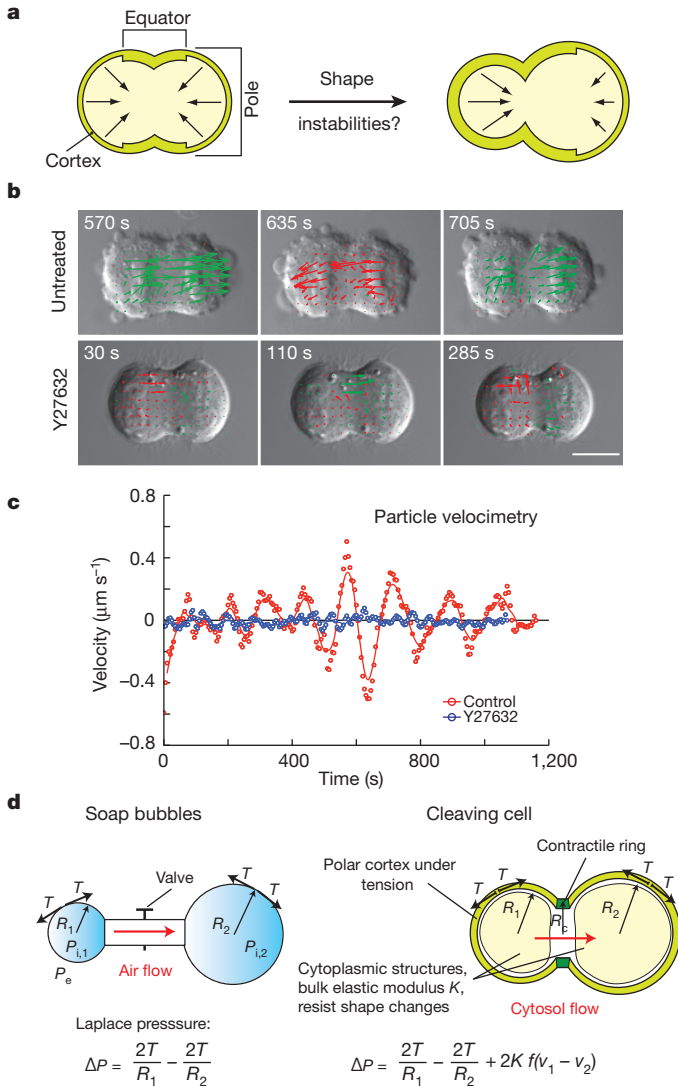


Figure 1 | The shape of the cleaving cell displays inherent shape instabilities. **a**, Polar cortex contraction could destabilize the cleavage furrow position. **b**, Stills from differential interference contrast (DIC) time-lapses of cytokinesis in control and Y27632-treated cells. Green and red arrows, rightward and leftward particle image velocimetry flow fields respectively; scale bar, 10 μm . **c**, Mean velocity of the flow field projected on the long cell axis as a function of time in the control (red circles) and the Y27632-treated (blue circles) cells pictured in **b**. Solid lines, data smoothed by factor 5. **d**, Left, schematic of the experiment demonstrating the instability of two soap bubbles of interfacial tension T connected by a pipe. P_e , outside pressure; $P_{i,1}$, $P_{i,2}$, pressures inside the bubbles; R_1 , R_2 , bubble radii. Right, schematic of a cytokinetic cell. R_1 , R_2 , radii of the future daughter cells; R_c , cleavage furrow radius; T , polar cortex tension; $Kf(v_1 - v_2)$, cell elastic response; ΔP is the pressure difference between the poles.

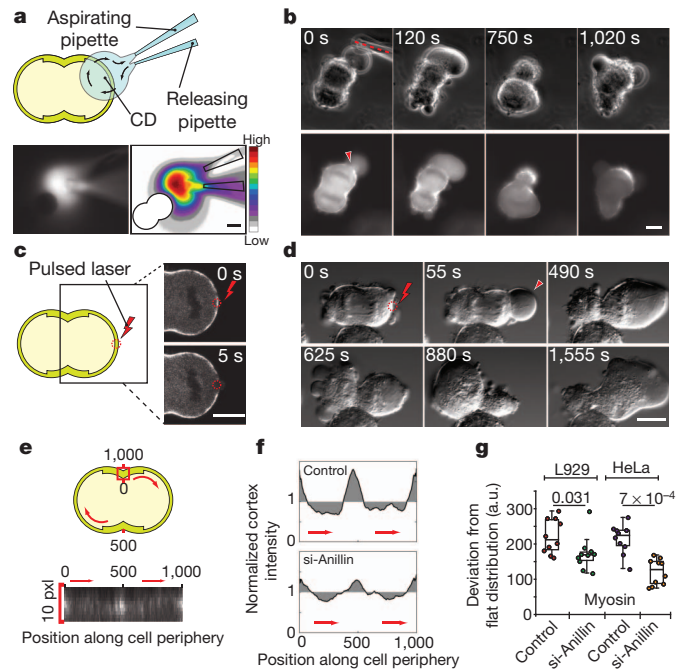


Figure 2 | Perturbation of the polar cortex enhances cytokinetic shape instabilities. **a**, Top, the local drug delivery experiment; bottom, fluorescent image of the released solution (visualized with Alexa Fluor 568; left, greyscale; right, pseudo-colour). **b**, Montage of a cleaving HeLa cell expressing actin-GFP, treated with cytochalasin D (CD) at a pole. 0 s, onset of cytochalasin D application. Red dotted line, position of the releasing pipette; the red arrowhead points to the rupture in the actin cortex. **c**, Left, the laser ablation experiment; right, example of polar cortex ablation in a HeLa cell expressing MRLC-tdRFP (red circle, ablation spot). **d**, Time-lapse of shape oscillations in a cleaving HeLa cell ablated at the polar cortex (red circle). 0 s, onset of ablation. Arrowhead points to the bleb induced by cortex ablation. Scale bars, 10 μm . **e**, Left, description of the measurement of cortex distribution along the cell periphery. **f**, Representative plots of the cortical distribution of MRLC-tdRFP in a control and an anillin-depleted cell (see Supplementary Fig. 1 and Methods). **g**, Integrals of cortical myosin distributions (shaded regions in **f**), in control and anillin-depleted cells shortly before the onset of oscillations. Boxes extend from the 25th to 75th percentiles, with a line at the median. P values, Mann-Whitney U test ($n = 10$). a.u., arbitrary units.

MRLC oscillated in phase opposition with the size of the poles, indicating that cortex density increases as a pole contracts (Fig. 3a, b, Supplementary Fig. 8 and Supplementary Movies 8 and 9). We thus refined our model to include the effect of cortex turnover. With τ the turnover time, cortex dynamics are described by $d(Sc)/dt = S(c_0 - c)/\tau$, where c is the cortex density, c_0 the steady-state density and S the surface area of a pole. This equation implies that when perturbed, cortex density and the resulting tension relax to their reference levels on a timescale τ (Supplementary Model section 3). Cytoplasm movements are described by $2\alpha dv_i/dt = -\Delta P$ where α is a friction coefficient limiting the variation of the normalized volume v_i of a pole in response to a pressure difference ΔP between the poles (Fig. 1d). For intermediate values of the stability parameter T_0/R_0K and for slow enough actin turnover, the system described by these two equations exhibits spontaneous shape oscillations (Fig. 3c, d and Supplementary Model). The underlying mechanism is as follows: if polar contraction is faster than cortex turnover, the cortex accumulates during contraction. As elasticity slows down contraction, cortex turnover leads to tension decrease at the contracting pole and increase at the opposing pole, and reversal of the oscillation direction. Cortex and pole size dynamics during oscillations could be very accurately fitted to the driving equations of the model (Fig. 3e, Supplementary Fig. 9 and Model Section 4). These fits allowed us to extract mechanical parameters, including the stability parameter T_0/R_0K from the data. We could thus position the oscillating cells in the state diagram and found that they were located in the linearly unstable oscillatory region (Fig. 3d and Supplementary Fig. 9). The period of the simulated oscillations was in good agreement with the period measured experimentally (Fig. 3c and Supplementary Fig. 8). Finally, the model predicts that the cell has a characteristic response time to pressure changes given by the ratio of friction to elasticity $\tau_c = \alpha/2K$. Consistently, we observed that cortex intensities reached a maximum and started decreasing before the reversal of the oscillation direction (Fig. 3b and Supplementary Fig. 8c, e). These observations show that the model (summarized in Fig. 4a) is in very good agreement with the experimental data.

Our findings indicate that polar tension must be tightly controlled during cytokinesis, raising the question of how this control is achieved. One striking difference between oscillating and control cells is that while control cells display blebs at both poles throughout division (Supplementary Movie 1), in oscillating cells, blebbing seems to alternate between the poles (Supplementary Movies 8 and 9). As bleb growth releases cortical tension and the resulting intracellular pressure^{18,25}, it is likely that blebbing effectively lowers polar tension and thus limits polar contractions (Fig. 4a). We tested the effect of inducing blebbing by laser ablation in oscillating anillin-depleted HeLa cells. Bleb induction in a contracting pole stopped the expansion of the opposite pole, while cytoplasm flowed into the bleb (Fig. 4b and Supplementary Movie 10), indicating that bleb formation can interfere with polar contraction. We then globally inhibited blebbing during cell division by treating cells with wheat germ agglutinin (WGA) and concavalin A, tetraivalent lectins that crosslink and stabilize the cortex from the outside²⁶. These treatments strongly decreased bleb formation during cytokinesis and triggered shape oscillations and division failure in 35% (WGA, $n = 102$) to 45% (conavalin A, $n = 130$) of L929 cells (Fig. 4c and Supplementary Movie 11), suggesting that hindering bleb formation favours shape instabilities. On the basis of these experiments, we propose that during cytokinesis, blebs act as pressure sinks that reduce contractile tension and the resulting intracellular pressure at the cell poles (Fig. 4a). Importantly, bleb size must be tightly controlled, as the formation of very large blebs can destabilize cell shape by introducing, rather than buffering, a tension imbalance between the poles (Fig. 2b, d and Supplementary Discussion).

Notably, shape instabilities arise only for high polar tensions (Fig. 3d), and high tension also favours bleb formation. Cortical tension and bleb formation could thus constitute a self-regulating mechanical feedback loop. Bleb nucleation results either from local detachment of

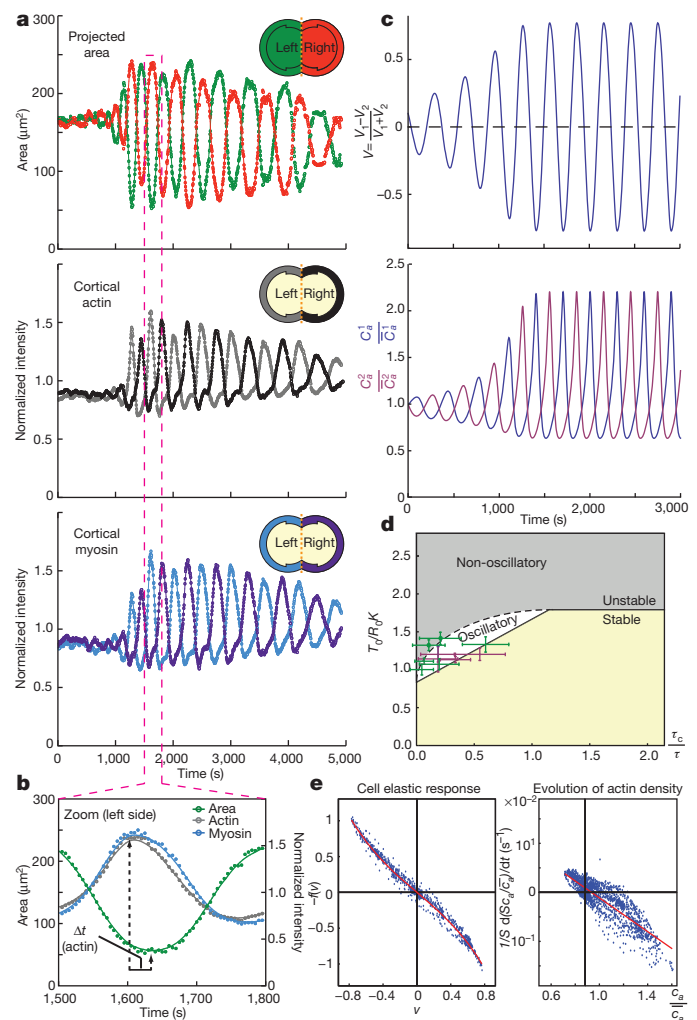


Figure 3 | Dynamic characterization and model of shape oscillations.

a, Cross-sectional areas of the two poles of an oscillating anillin-depleted L929 cell and mean intensities of cortical actin and myosin, left and right of the cleavage furrow, as a function of time. **b**, Zoom on the regions highlighted in **a** for the left pole. **a**, **b**, Intensities are normalized to their average. Circles, data points; solid lines, data smoothed by factor 5. **c**, Simulation of the evolution of the difference of volumes v of the two poles (top) and of the polar actin densities (bottom: c_a , cortex density; \bar{c}_a , mean cortex density) using parameters extracted from the cell in **a** (see Supplementary Model). **d**, Phase diagram of the linear stability analysis of a symmetric cell with respect to volume perturbation in the model coupling actin turnover and cell mechanics, in the parameter space (τ_c/τ , T_0/R_0K), with $R_c/R_0 = 0.8$; T_0 is the polar tension before oscillations. Cells are positioned on the diagram according to parameters extracted from fits as in **e** (magenta, anillin depleted cells; green, laser-ablated cells). Error bars, standard deviation of the fit parameters. **e**, Fit of the experimental parameters of the cell in **a** (blue points) to the driving equations of the model (red curves). Left, fit to the dynamic equation of v , yielding the elastic response of the cell $f(v)$; right, fit to the equation for cortex density (see Supplementary Model).

the membrane from the cortex or from local cortex tearing (Supplementary Fig. 10). Both mechanisms are favoured by high cortical tension⁵. It is thus possible that the cytokinetic cortex is tuned to form blebs, and thus release tension, as soon as polar tension exceeds the threshold for shape instabilities. Alternatively, cytokinetic bleb formation could result from a mechanosensing signalling cascade. However, we found that calcium ions, common mediators of mechanosensory responses, which have been shown to be implicated in blebbing²⁷, were not involved in cytokinetic bleb formation and shape oscillations (Supplementary Fig. 11). Other elements such as the spindle also contribute to maintaining the cleavage furrow at the equator^{20,21}, and mechanical properties, such as cell elasticity, can also prevent shape

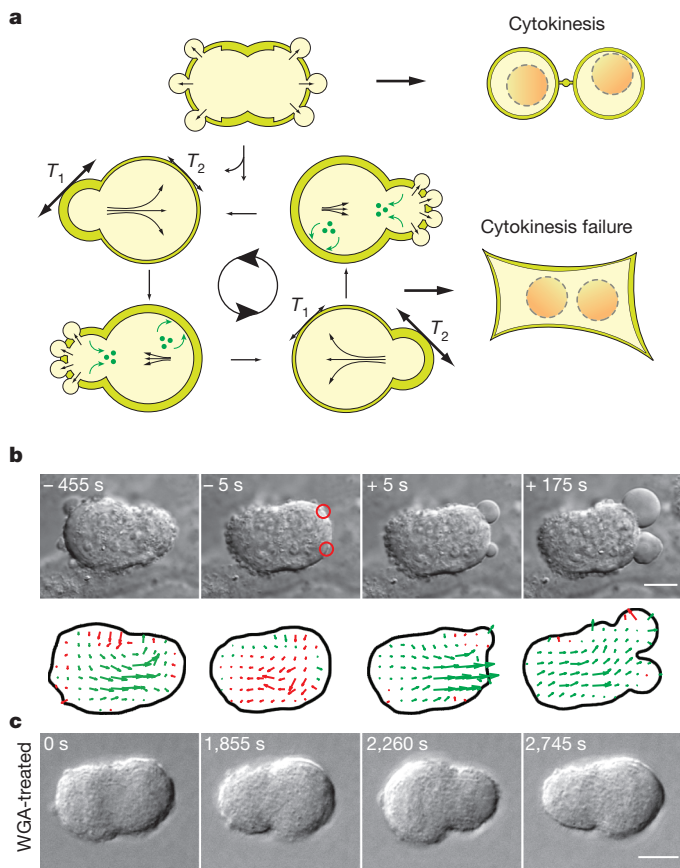


Figure 4 | Cell shape stability during cytokinesis. **a**, Top, in control cells, bleb formation constantly releases polar pressure and helps ensure the success of cytokinesis. Bottom, high polar tension or a large pressure imbalance between the poles can trigger shape oscillations. Model of the oscillations: if cortex contraction (resulting from tension T) is faster than cortex turnover (green arrows), cortex contractility, dynamics and cell elasticity result in cycles of cortex contraction, disassembly and subsequent reassembly. **b**, Laser ablation at the contracting pole of an oscillating anillin-depleted HeLa cell arrests the flow of cytoplasm between the poles and causes a premature reversal of the flow direction (15 cells out of 16). Top, DIC time-lapse; bottom, corresponding particle image velocimetry vectors and cell outline. Two consecutive ablations were performed (red circles); 0 s, first frame after ablation. **c**, DIC image series of a cytokinetic L929 cell oscillating after WGA treatment. Scale bars, 10 μm .

instabilities. Therefore, polar bleb formation may act as a partly redundant safety mechanism. Such redundancy is common in cytokinesis and is thought to provide robustness to the process²⁸. Future studies will have to elucidate how exactly cellular mechanical properties, molecular pathways and bleb formation are integrated into a fine-tuned mechanical system ensuring the success of cytokinesis.

Our findings may have important implications for asymmetric cell division. A recent study in *C. elegans* QR.a neuroblasts¹¹ shows that asymmetric divisions can be the direct result of myosin-driven cortex contraction of one of the poles at the expense of the other, a mechanism that could also be relevant for asymmetric division of *Drosophila* neuroblasts⁸ (see Supplementary Discussion). An important open question is how such an asymmetric shape, with more cortical myosin and thus a higher tension in the smaller pole, can be stable. It will be interesting to investigate the mechanics of contractility-driven asymmetric divisions.

Finally, the oscillation mechanism proposed here could be relevant for other systems where cortex oscillations are observed, such as during epithelial contractions in *Drosophila* development^{29,30} (Supplementary Discussion). A common feature of all these oscillations is that they display cycles of cortex contraction, disassembly and reassembly,

suggesting they might be driven by a common mechanism coupling contractility, cortex turnover and cell elasticity.

METHODS SUMMARY

Culture conditions and cell treatments. Cell culture, drug treatments, gene silencing by RNA interference, knockdown quantification by quantitative PCR and cell transfections are detailed in the Supplementary Methods.

Local cytochalasin D delivery. We conducted local delivery of cytochalasin D using two micropipettes mounted on a motorized micromanipulator (Eppendorf) connected to pressure regulators. Imaging was performed on a Carl Zeiss Axiovert 200M inverted microscope.

Local laser ablation. Laser ablation experiments were performed as described in ref. 18 on a scanning confocal microscope (Olympus FV1000) equipped with two scanning heads. For ablation, we used a 405-nm picosecond-pulsed laser with a nominal power of 3 mW for 2.5 s.

Particle image velocimetry and image processing. Cytoplasmic flows in dividing cells were determined with a custom-made particle image velocimetry algorithm in MATLAB (MathWorks). Images were processed, automatically analysed and visualized using custom code in MATLAB (MathWorks) (detailed in the Supplementary Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 December 2010; accepted 13 June 2011.

Published online 7 August 2011.

- Eggert, U. S., Mitchison, T. J. & Field, C. M. Animal cytokinesis: from parts list to mechanisms. *Annu. Rev. Biochem.* **75**, 543–566 (2006).
- Robinson, D. N. & Spudich, J. A. Mechanics and regulation of cytokinesis. *Curr. Opin. Cell Biol.* **16**, 182–188 (2004).
- Pollard, T. D. Mechanics of cytokinesis in eukaryotes. *Curr. Opin. Cell Biol.* **22**, 50–56 (2010).
- Wang, Y. L. The mechanism of cortical ingression during early cytokinesis: thinking beyond the contractile ring hypothesis. *Trends Cell Biol.* **15**, 581–588 (2005).
- Charras, G. & Paluch, E. Blebs lead the way: how to migrate without lamellipodia. *Nature Rev. Mol. Cell Biol.* **9**, 730–736 (2008).
- Zhang, W. & Robinson, D. N. Balance of actively generated contractile and resistive forces controls cytokinesis dynamics. *Proc. Natl Acad. Sci. USA* **102**, 7186–7191 (2005).
- Werner, M., Munro, E. & Glotzer, M. Astral signals spatially bias cortical myosin recruitment to break symmetry and promote cytokinesis. *Curr. Biol.* **17**, 1286–1297 (2007).
- Cabernard, C., Prehoda, K. E. & Doe, C. Q. A spindle-independent cleavage furrow positioning pathway. *Nature* **467**, 91–94 (2010).
- Yoneda, M. & Dan, K. Tension at the surface of the dividing sea-urchin egg. *J. Exp. Biol.* **57**, 575–587 (1972).
- Effler, J. C., Iglesias, P. A. & Robinson, D. N. A mechanosensory system controls cell shape changes during mitosis. *Cell Cycle* **6**, 30–35 (2007).
- Ou, G., Stuurman, N., D'Ambrosio, M. & Vale, R. D. Polarized myosin produces unequal-size daughters during asymmetric cell division. *Science* **330**, 677–680 (2010).
- Straight, A. F., Field, C. M. & Mitchison, T. J. Anillin binds nonmuscle myosin II and regulates the contractile ring. *Mol. Biol. Cell* **16**, 193–201 (2005).
- Hickson, G. R. & O'Farrell, P. H. Rho-dependent control of anillin behavior during cytokinesis. *J. Cell Biol.* **180**, 285–294 (2008).
- Piekny, A. J. & Glotzer, M. Anillin is a scaffold protein that links RhoA, actin, and myosin during cytokinesis. *Curr. Biol.* **18**, 30–36 (2008).
- Smith, T. C., Fang, Z. & Luna, E. J. Novel interactors and a role for supervillin in early cytokinesis. *Cytoskeleton* **67**, 346–364 (2010).
- Dean, S. O., Rogers, S. L., Stuurman, N., Vale, R. D. & Spudich, J. A. Distinct pathways control recruitment and maintenance of myosin II at the cleavage furrow during cytokinesis. *Proc. Natl Acad. Sci. USA* **102**, 13473–13478 (2005).
- Rankin, K. E. & Wordeman, L. Long astral microtubules uncouple mitotic spindles from the cytokinetic furrow. *J. Cell Biol.* **190**, 35–43 (2010).
- Tinevez, J. Y. et al. Role of cortical tension in bleb growth. *Proc. Natl Acad. Sci. USA* **106**, 18581–18586 (2009).
- Kosako, H. et al. Rho-kinase/ROCK is involved in cytokinesis through the phosphorylation of myosin light chain and not ezrin/radixin/moesin proteins at the cleavage furrow. *Oncogene* **19**, 6059–6064 (2000).
- Bringmann, H. & Hyman, A. A. A cytokinesis furrow is positioned by two consecutive signals. *Nature* **436**, 731–734 (2005).
- Werner, M. & Glotzer, M. Control of cortical contractility during cytokinesis. *Biochem. Soc. Trans.* **36**, 371–377 (2008).
- Plateau, J. A. F. *Statique Expérimentale et Théorique des Liquides Soumis aux Seules Forces Moléculaires*, Vol. 2, Ch. 7 (Gauthier-Villars, 1873).
- Mitchison, T. J., Charras, G. T. & Mahadevan, L. Implications of a poroelastic cytoplasm for the dynamics of animal cell shape. *Semin. Cell Dev. Biol.* **19**, 215–223 (2008).
- Riedl, J. et al. Lifeact: a versatile marker to visualize F-actin. *Nature Methods* **5**, 605–607 (2008).

25. Charras, G. T., Yarrow, J. C., Horton, M. A., Mahadevan, L. & Mitchison, T. J. Non-equilibration of hydrostatic pressure in blebbing cells. *Nature* **435**, 365–369 (2005).
26. Rosenblatt, J., Cramer, L. P., Baum, B. & McGee, K. M. Myosin II-dependent cortical movement is required for centrosome separation and positioning during mitotic spindle assembly. *Cell* **117**, 361–372 (2004).
27. Blaser, H. *et al.* Migration of zebrafish primordial germ cells: a role for myosin contraction and cytoplasmic flow. *Dev. Cell* **11**, 613–627 (2006).
28. Burgess, D. R. & Chang, F. Site selection for the cleavage furrow at cytokinesis. *Trends Cell Biol.* **15**, 156–162 (2005).
29. Blanchard, G. B., Murugesu, S., Adams, R. J., Martinez-Arias, A. & Gorfinkiel, N. Cytoskeletal dynamics and supracellular organisation of cell shape fluctuations during dorsal closure. *Development* **137**, 2743–2752 (2010).
30. Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin–myosin network drive apical constriction. *Nature* **457**, 495–499 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. S. Bois, A. G. Clark, S. W. Grill, C. P. Heisenberg, J. Howard, A. A. Hyman, J. F. Joanny, D. K. Lubensky, A. Oates, M. Piel, I. M. Tolic-Norrelykke, W. Zachariae and M. Zerial for discussions and comments on the manuscript, and J. Roensch and the Max Planck Institute of Molecular Cell Biology and Genetics Light Microscopy Facility for technical assistance. This work was supported by the Polish Ministry of Science and Higher Education and by the Max Planck Society.

Author Contributions J.S., M.B., G.S. and E.P. designed the research and wrote the paper; J.S. performed the experiments except the local drug delivery; M.B. developed the image analysis tools; J.S., M.B. and G.S. analysed the data; A.O. and J.-Y.T. designed and performed the local drug-delivery experiments; G.S. developed the theoretical model.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.P. (paluch@mpi-cbg.de).

METHODS

Cell lines and culture. L929 and HeLa fibroblasts were grown in DMEM (GIBCO, Invitrogen) supplemented with 10% foetal bovine serum, 1% glutamine and 1% penicillin-streptomycin in polystyrene culture flasks (Nunc). For all the time-lapse experiments, the cells were plated on glass-bottom dishes (Mattek), maintained at 37 °C and supplied with 5% CO₂ on the microscope stage. All cell lines are adherent and maintain attachments to the substrate throughout division (Supplementary Fig. 5 and Supplementary Movie 5). The stable HeLa-GFP-actin line was a gift from F. Buchholz. The stable HeLa-EB3-GFP-Lifect-mCherry line was a gift from the laboratory of M. Piel.

Cell treatments, plasmids, RNAi and cell transfection. Rho-kinase was inhibited with Y27632 (Merck Bioscience, at least 30 min treatment at 10 µM). To stabilize the cortex and limit bleb formation, we used wheat germ agglutinin (Invitrogen) and concavalin A (Sigma Aldrich-Chemie) at 25 µg ml⁻¹ for up to 2 h as described in refs 26 and 31. Lifect-GFP²⁴ was a gift from R. Wedlich-Söldner. MRLC-tdRFP was a gift from G. Charras. mCherry-CAAX was generated by replacing the GFP cassette with an mCherry cassette in a CAAX-GFP plasmid (gift from G. Charras), by cutting with BamHI and EcoRI (performed by M. Bergert). For experiments requiring increased substrate adhesiveness, glass-bottom dishes were coated with 15 µg ml⁻¹ collagen type I from rat tail (Sigma) and fibronectin 0.1% solution (diluted 1:1,000 in H₂O) from bovine plasma (Sigma). Mouse stealth anillin short interfering RNA (siRNA) (5'-CCGUUGGAAGGUCACAUUGUUUAA), stealth Gelsolin (5'-ACGGGUGAUGCCUAU GUCAUCCUAA) and scrambled control were obtained from Invitrogen. Mouse cofilin siRNA (5'-GAAUCAAGCAUGAAUUACATT) was obtained from Ambion. Human anillin siRNA (5'-GAUGCUUGCUACAAACCUATT) was obtained from Qiagen. For L929 cells, transfection of 2 µg per well plasmid DNA or 4 pmol per well siRNA was achieved by electroporation using the Cell Line 96-well Nucleofector Kit SE (Lonza) according to the instructions supplied (program code EH-100). For siRNA transfection in HeLa cells, Lipofectamine RNAiMax (Invitrogen) was used. The final concentration of siRNA was 10 nM. For co-transfection of plasmid and siRNA in HeLa cells, Lipofectamine 2000 (Invitrogen) was used. Final concentrations were 10 nM for siRNA and 0.2 µg ml⁻¹ for plasmid. Cells were used 30 h after transfection in laser ablation and anillin depletion experiments. Intracellular Ca²⁺ was chelated by an incubation of no less than 15 min in 40 µM BAPTA-AM (Invitrogen). Extracellular Ca²⁺ was removed by performing the experiments in Ca²⁺-free normal Tyrode's solution. Additional chelation with 500 µM EGTA (Sigma) was tested but had no further effect. Microtubules were depolymerized with 5 µM nocodazole (Sigma).

RNA isolation and quantitative PCR. The extent of anillin depletion in L929 and HeLa cells (Supplementary Fig. 14) was estimated by performing qPCR with reverse transcription and comparing the levels of messenger RNA (mRNA) to endogenous levels of mBeta-2- microglobulin (L929) or GAPDH (HeLa) mRNA. Total RNA was isolated using the RNeasy kit (Qiagen) according to the manufacturer's instructions. RNA was cleaned of genomic DNA by DNase digestion with RNase-Free DNase Set digest kit (Qiagen). Complementary DNA was subsequently synthesized from 0.5 µg of RNA using SuperScript III RNase H-Reverse Transcriptase kit (Invitrogen). Comparative qPCR was performed using the Absolute qPCR SYBR Green mix (Thermo Fisher Scientific) and an Mx3000P cyclo (Stratagene). Primers were as follows: L929 Anillin, forward 5'-AACTCGAATGCAAAGGCTTG, reverse 5'-AATGGGTGGCTTAGAAGGTG; L929 mBeta-2-microglobulin, forward 5'-ATTCACCCCACTGAGACTG, reverse 5'-TGCTATTCTTTCTGCGTGC; HeLa Anillin, forward 5'-GACAAATCTACGACACAGGAG, reverse 5'-CTTGGATGGCCTTTGTATTGG; HeLa GAPDH, forward 5'-GAGTCAACGGATTTGGTCGT, reverse 5'-TGGAAGATGGTGATGGGATT.

Local cytochalasin D delivery. Local delivery of cytochalasin D³² was performed on a Carl Zeiss Axiovert 200M inverted microscope equipped with motorized micromanipulator arms (Eppendorf TransferMan NK2) and imaged through a Carl Zeiss Neofluar ×63 (numerical aperture = 0.75) objective. Micropipettes were pulled with a flaming pipette-puller (P-97, Sutter Instruments), and forged with a home-built forge. The releasing micropipette was loaded with 20 µl of 5 µg ml⁻¹ cytochalasin D (Sigma) diluted in culture medium added with Alexa Fluor 568 Succinimidyl Ester diluted to a final concentration of 600 nM (Invitrogen), and mounted together with the aspiration micropipette on a double pipette holder (Eppendorf TwinTip-Holder). The aspiration pipette was positioned slightly behind the releasing pipette (Fig. 2a). Both micropipettes were connected to pressure regulators: a Femtojet Microinjector (Eppendorf) for the releasing pipette and a pump for the aspiration pipette (Eppendorf CellTram Air). The flowfield was monitored by tracking fluorescent Alexa Fluor 568. The pipettes were positioned so that only one pole of a dividing cell was exposed to the released drug. The release of cytochalasin D was maintained only until local bleb formation

was observed (about 30 s). For the dimethylsulphoxide control, cytochalasin D was replaced by dimethylsulphoxide at 1:1,000 dilution in the releasing pipette.

Laser ablation and cell imaging. Laser ablation experiments were performed as described in ref. 18 on a scanning confocal microscope (Olympus FV1000) equipped with two scanning heads. The first was used for standard imaging; the second was coupled to a laser head (LDH-P-C-405B, PicoQuant), driven by a power source (PDL 800-B, PicoQuant) delivering 405-nm picosecond pulses with a nominal power of 3 mW. For ablation, the pulsed laser underwent spiralling movements within a circle of 800-nm nominal diameter for 2.5 s. Unless otherwise specified, cells were imaged transversally through a single central confocal Z plane, midway between the substrate and the top of the cell, through an Olympus UPlanSApo ×60 oil immersion objective (numerical aperture = 1.35). Imaging for volume measurements was performed on an Andor IQ spinning disc confocal microscope by acquiring Z stacks through the entire volume of cells, through an Olympus UPlanSApo ×60 water immersion objective (numerical aperture = 1.2) and ×1.6 Optovar. Long-term time-lapse microscopy was performed on an Olympus-IX81 widefield microscope through an Olympus UPlanSApo ×20 air objective (numerical aperture = 0.5).

Image processing and automated image analysis. Images were processed, automatically analysed and visualized using custom code in MATLAB (MathWorks). Particle image velocimetry³³ was performed on DIC images yielding local measures of the cytoplasmic flows. Particle image velocimetry is an image-processing-based technique for measuring fluid displacement by cross-correlation between images from consecutive time points. For illustration purposes, the flow vectors were colour-coded according to their angles with respect to the cleavage furrow. Vectors with rightward direction were coloured green, and vectors with leftward direction were coloured red. The mean vector component along the pole-pole axis was calculated for each time-point then smoothed with a moving average of window size 5. The angular distribution and amplitudes of the mean vectors per time-point formed the basis of the classification of cells as either displaying cytoplasmic flows or not (Supplementary Fig. 2). Flows in cytokinetic cells were considered significant when their amplitude was more than two times higher than the amplitude of flows typically observed in metaphase cells.

Cell contour detection and definition of cellular regions: segmentation and tracking of cells was based on a combination of image filtering, ellipse fitting, adaptive thresholding and edge detection operations on raw fluorescent intensities from one of the fluorescent channels. The cortex was defined as a region of fixed depth underlying the segmented cell contour. The cytoplasmic region was taken as the difference between the segmented cell area and the cortical region. The cleavage furrow was defined by a straight line between two points, the positions of which were determined automatically along the cell contour. The end points were repositioned manually when the algorithm failed to position the furrow line correctly. The furrow line delimited the left side of the cell from the right side. The definition of the cell contour, cortical, cytoplasmic and left/right delimitations allowed for the extraction of region-specific morphological parameters, intensity values and oscillation periods (Supplementary Fig. 12).

Cortex intensity measurements: mean fluorescence intensities I_{mean} of F-actin (Lifect) and myosin (MRLC) in the cortex were determined in the defined cortical region (described above). As the depth of this region was kept constant while the apparent thickness of the cortex varied both along the cell contour and within a time-lapse, the cortical region was chosen so as to overestimate the actual cortex. The region therefore included pixels belonging both to a varying depth of cortex and to the cytoplasm. The mean intensities measured in this region thus constitute a combined measure of both the average cortical intensity per unit area and the actual thickness of the cortex. Cortical intensity profiles (Fig. 2f and Supplementary Figs 1, 6 and 7) depict the mean cortical intensity along the cell contour $I_{\text{mean}}(s)$, where s is the position along the cell contour. The profiles start at one intersection of the furrow line with the cell contour and then follow the cell contour in the cortex region in a clockwise direction. To allow for comparison of cells with different contour lengths, the path length was set to a constant length of 1,000 points and intensity values were interpolated by a factor of 1,000 per actual path length. The intensity values were then normalized to the global cortex mean intensity ($I_{\text{mean}}(s)/\langle I_{\text{mean}} \rangle$), so that values greater (respectively smaller) than 1 indicate that the local cortex intensity is larger (respectively smaller) than the global cortex mean. To measure the deviation from a perfectly flat intensity profile, which would indicate a perfectly uniform distribution of intensities in the cortex, we integrated the deviations of the intensity profiles of cortical proteins, normalized as described above, from $1 (|I_{\text{mean}}(s)/\langle I_{\text{mean}} \rangle - 1| \text{ ds})$.

Volume measurements. Volume measurements during cytokinesis were performed on dividing L929 cells (Supplementary Fig. 13). Oscillations were induced by anillin depletion, as described above. Cells were loaded with an intracellular fluorescent dye (2.5 µM CMFDA (5-chloromethylfluorescein diacetate), Invitrogen) excitable at 488 nm and emitting at approximately 520 nm for 30 min at 37 °C and 5% CO₂

and then changed into conditioned medium for 1 h before imaging. 5 μ M Alexa Fluor 568 Succinimidyl Ester (Invitrogen) was then added to the medium as an extracellular dye excitable at 561 nm and emitting at approximately 590 nm. Imaging and data acquisition were performed on an Andor IQ spinning disc confocal microscope. Time-lapses were bleach corrected using a custom plugin, the Alexa Fluor 568 channel inverted and multiplied with the CMFDA channel using Fiji Is Just ImageJ. Surface rendering and volume calculations were then performed in Imaris $\times 64$ (Bitplane). We estimate an error in volume measurements no greater than 10.65%, based on an allowed error of 1 voxel diagonal in

determining the cell boundary at every point of a spherical cell of 3,000 μm^3 (imaging voxels were on average 181 nm \times 181 nm \times 580 nm).

31. Kunda, P., Pelling, A. E., Liu, T. & Baum, B. Moesin controls cortical rigidity, cell rounding, and spindle morphogenesis during mitosis. *Curr. Biol.* **18**, 91–101 (2008).
32. O'Connell, C. B., Warner, A. K. & Wang, Y.-I. Distinct roles of the equatorial and polar cortices in the cleavage of adherent cells. *Curr. Biol.* **11**, 702–707 (2001).
33. Raffel, M., Willert, C. E., Wereley, S. T. & Kompenhans, J. *Particle Image Velocimetry – A Practical Guide* (Springer, 2007).

A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation

Vlad C. Seitan^{1,2,3*}, Bingtao Hao^{4*}, Kikuë Tachibana-Konwalski^{5*}, Thais Lavagnoli^{1,3}, Hegias Mira-Bontenbal^{1,3}, Karen E. Brown^{1,3}, Grace Teng⁶, Tom Carroll³, Anna Terry^{1,3}, Katie Horan⁷, Hendrik Marks⁸, David J. Adams⁹, David G. Schatz^{6,10}, Luis Aragon^{2,3}, Amanda G. Fisher^{1,3}, Michael S. Krangel⁴, Kim Nasmyth⁵ & Matthias Merkenschlager^{1,3}

Cohesin enables post-replicative DNA repair and chromosome segregation by holding sister chromatids together from the time of DNA replication in S phase until mitosis¹. There is growing evidence that cohesin also forms long-range chromosomal cis-interactions^{2–4} and may regulate gene expression^{2–10} in association with CTCF^{8,9}, mediator⁴ or tissue-specific transcription factors¹⁰. Human cohesinopathies such as Cornelia de Lange syndrome are thought to result from impaired non-canonical cohesin functions⁷, but a clear distinction between the cell-division-related and cell-division-independent functions of cohesin—as exemplified in *Drosophila*^{11–13}—has not been demonstrated in vertebrate systems. To address this, here we deleted the cohesin locus *Rad21* in mouse thymocytes at a time in development when these cells stop cycling and rearrange their T-cell receptor (TCR) α locus (*Tcra*). *Rad21*-deficient thymocytes had a normal lifespan and retained the ability to differentiate, albeit with reduced efficiency. Loss of *Rad21* led to defective chromatin architecture at the *Tcra* locus, where cohesin-binding sites flank the TEA promoter and the Ea enhancer, and demarcate *Tcra* from interspersed *Tcrd* elements and neighbouring housekeeping genes. Cohesin was required for long-range promoter-enhancer interactions, *Tcra* transcription, H3K4me3 histone modifications that recruit the recombination machinery^{14,15} and *Tcra* rearrangement. Provision of pre-rearranged TCR transgenes largely rescued thymocyte differentiation, demonstrating that among thousands of potential target genes across the genome^{4,8–10}, defective *Tcra* rearrangement was limiting for the differentiation of cohesin-deficient thymocytes. These findings firmly establish a cell-division-independent role for cohesin in *Tcra* locus rearrangement and provide a comprehensive account of the mechanisms by which cohesin enables cellular differentiation in a well-characterized mammalian system.

The somatic rearrangement of lymphocyte receptor loci is central to adaptive immunity¹⁶. Gene segments distributed over millions of base pairs of genomic DNA are transcribed, brought into proximity with each other, and recombined in a cell-lineage- and developmental-stage-specific fashion^{17–19}. In developing thymocytes, proliferation and differentiation are tightly linked and the activity of Rag (recombination activating gene) proteins is restricted to the G1 phase of the cell cycle²⁰. Early thymocytes at CD4[–] CD8[–] double-negative (DN) stages 1 and 2 proliferate in response to cytokines and briefly arrest at the DN3 stage, where they rearrange the TCR β locus (Fig. 1a). Pre-TCR signals drive a phase of proliferation that extends to the early CD4⁺ CD8⁺ double-positive (DP) stage. Shortly after the acquisition of CD4 and CD8, DP thymocytes lose the expression of the transferrin receptor CD71 (ref. 21) and become small, non-proliferating CD71[–]

DP cells (Fig. 1a), which represent the great majority of thymocytes. During their life span of 3 to 4 days, DP thymocytes undergo several rounds of *Tcra* rearrangement^{16,18,19}. Successful TCR expression and engagement selects a minority (3–5%) of DP thymocytes for differentiation via a CD4⁺ CD8^{lo} intermediate stage towards long-lived CD4 or CD8 single-positive (SP) cells, again with minimal proliferation^{16,22} (Fig. 1a).

To address the role of cohesin in *Tcra* rearrangement and thymocyte differentiation, we combined a conditional allele encoding the cohesin subunit Rad21 (*Rad21*^{lox}; Fig. 1b and Supplementary Fig. 1a) with a Cre transgene under the control of *Cd4* regulatory elements (CD4-Cre), which becomes active at the transition from the CD4[–] CD8[–] DN to the DP stage²³. Pilot experiments with yellow fluorescent protein (YFP) reporters showed CD4-Cre-dependent accumulation of YFP after the CD71⁺ proliferative stage in non-dividing DP thymocytes (Supplementary Fig. 1b). Proliferating CD4-Cre *Rad21*^{lox/lox} CD71⁺ DP cells showed partial locus deletion but retained >50% *Rad21* messenger RNA and protein (Fig. 1c). *Rad21* genomic deletion was essentially complete (>97%) and *Rad21* RNA and protein levels were substantially reduced in non-dividing DP thymocytes (Fig. 1c). Hence, cohesin was selectively depleted from non-dividing thymocytes. Importantly, CD4-Cre *Rad21*^{lox/lox} DP thymocyte numbers were normal (Fig. 1d). Intermediate CD4⁺ CD8^{lo} and mature CD4 SP and CD8 SP thymocytes accumulated slowly in CD4-Cre *Rad21*^{lox/lox} mice (Supplementary Fig. 2a) but were present in normal numbers by 6 weeks of age (Fig. 1d).

Bromodeoxyuridine (BrdU) incorporation into replicating DNA can identify proliferating thymocyte populations and track their differentiation²². Pulse-chase experiments labelled proliferating (CD71⁺) DP but not non-proliferating CD4⁺ CD8^{lo}, CD4 SP or CD8 SP thymocytes²² (Fig. 1e, 2 h time point). During the subsequent chase period, BrdU-labelled DP cells differentiated to become CD4⁺ CD8^{lo} and eventually CD4 or CD8 SP cells (Fig. 1e, left). This sequence of differentiation was preserved in CD4-Cre *Rad21*^{lox/lox} thymocytes, but the proportion of DP thymocytes that became CD4 SP or CD8 SP was reduced (Fig. 1e, right, and Supplementary Fig. 2b).

In continuous BrdU labelling experiments, the percentage of BrdU⁺ cells indicates population turnover²². Importantly, cohesin-deficient and control DP thymocytes labelled with similar kinetics (Fig. 1f). Consistent with the pulse-labelling data (Fig. 1e), the accumulation of CD4⁺ CD8^{lo} and CD4 SP subsets was reduced in CD4-Cre *Rad21*^{lox/lox} mice (Supplementary Fig. 2c). Hence, cohesin depletion impaired the differentiation of DP thymocytes, but not their survival.

Unlike many other differentiated cell types, mature thymocytes can be induced to re-enter the cell cycle. *In vitro* activated CD4-Cre

¹Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College London, Du Cane Road, London W12 0NN, UK. ²Cell Cycle Group, MRC Clinical Sciences Centre, Imperial College London, Du Cane Road, London W12 0NN, UK. ³Epigenetics Section, MRC Clinical Sciences Centre, Imperial College London, Du Cane Road, London W12 0NN, UK. ⁴Department of Immunology, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁵Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ⁶Department of Immunobiology, Yale University School of Medicine, 300 Cedar Street, New Haven, Connecticut 06520-8011, USA. ⁷Central Biological Services, Imperial College London, Du Cane Road, London W12 0NN, UK. ⁸Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, Radboud University, 6500 HB Nijmegen, The Netherlands. ⁹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. ¹⁰Howard Hughes Medical Institute, Yale University School of Medicine, 300 Cedar Street, New Haven, Connecticut 06520-8011, USA.

*These authors contributed equally to this work.

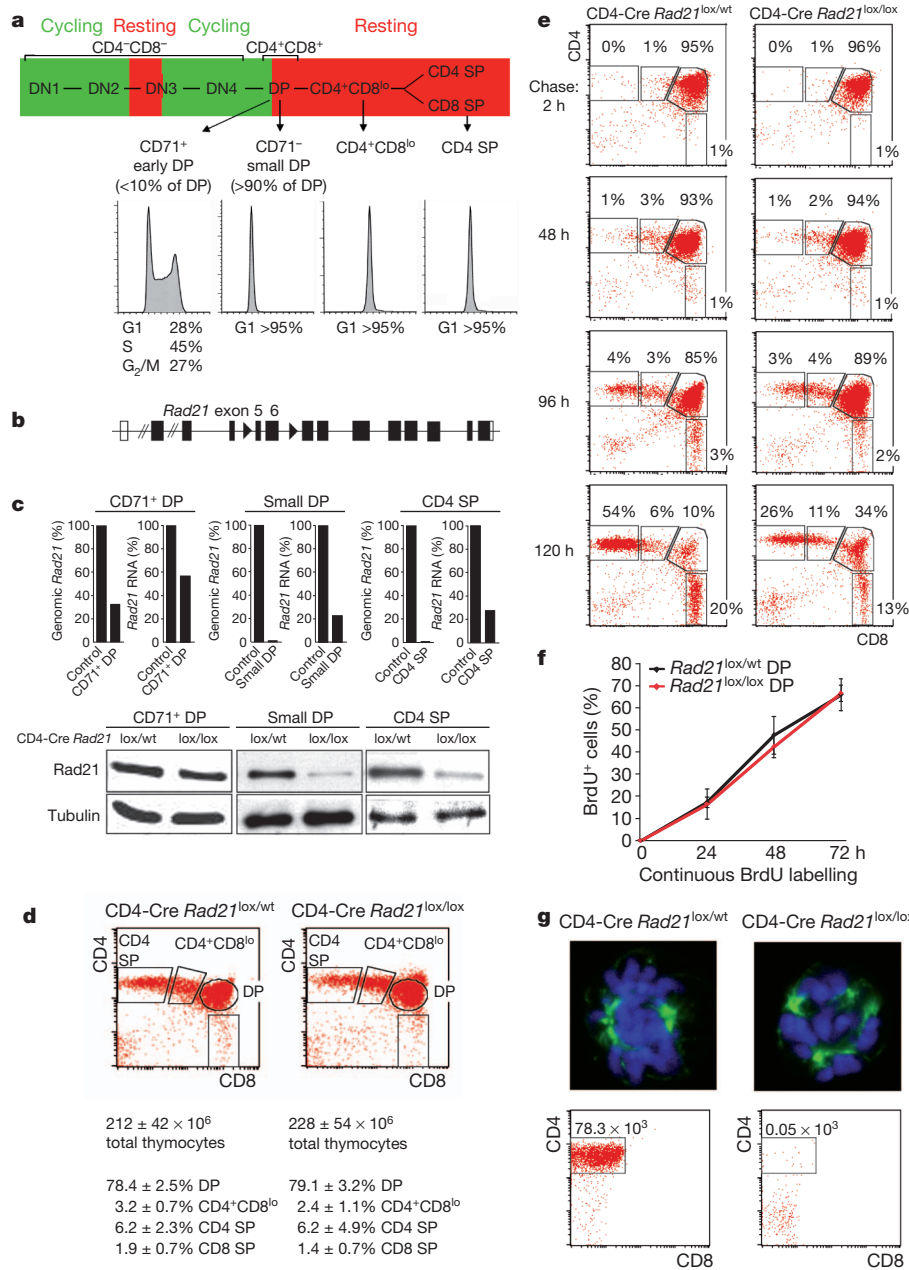


Figure 1 | Genetic cohesin depletion in non-dividing thymocytes.

a, Thymocyte differentiation from left to right: CD4⁻CD8⁻ DN stages 1 to 4; CD4⁺CD8^{lo} DP, CD4⁺CD8^{lo}; CD4 or CD8 SP cells. Proliferation is in green, cell cycle arrest in red. Histograms show DNA content. **b**, Conditional *Rad21* allele (see Supplementary Fig. 1a). **c**, Real-time genomic PCR of *Rad21* locus deletion, RT-PCR of *Rad21* RNA and western blotting of Rad21 protein. **d**, Cell numbers and flow cytometric analysis of thymocyte subsets in 6-week-old CD4-Cre *Rad21*^{lox/lox} and CD4-Cre *Rad21*^{lox/wt} mice (mean ± standard

Rad21^{lox/lox} CD4 SP thymocytes showed abnormal mitotic figures with multiple spindles, chromosome segregation defects (Fig. 1g and Supplementary Fig. 3) and poor survival (Fig. 1g). CD4-Cre-mediated deletion of *Rad21* therefore generates thymocytes that die when forced to divide, yet have a normal lifespan as non-dividing cells *in vivo*. This allows the interrogation of cohesin functions in interphase, independent of essential cohesin functions during cell division.

Rad21 chromatin immunoprecipitation and sequencing (ChIP-seq) mapped cohesin to key positions within the *Tcr* locus in DP thymocytes (Fig. 2a). Cohesin was abundant at the locus control region²⁴, which separates the *Tcr* enhancer *Eα* from the neighbouring *Dad1* housekeeping gene^{25,26}. Other prominent cohesin sites separated

deviation (s.d., *n* = 12). **e**, Pulse chase analysis of CD4-Cre *Rad21*^{lox/wt} and CD4-Cre *Rad21*^{lox/lox} thymocytes. Dot blots are gated on BrdU⁺ cells (see Supplementary Fig. 2b). **f**, Continuous BrdU labelling for DP thymocyte turnover. See Supplementary Fig. 2c for CD4⁺CD8^{lo} and CD4 SP subsets (mean ± s.d., *n* = 3–5 per data point). **g**, Top, metaphase spreads of 2-day activated thymocytes stained for α-tubulin (green) and DNA (DAPI, blue; see Supplementary Fig. 3). Magnification, ×2,500. Bottom, cells recovered after 5 days.

the *Tcr* TEA promoter from the *Tcrd* enhancer *Eδ*, which controls *Tcrd* gene segments that are interspersed within the *Tcr* locus but follow a distinct developmental stage-specific program¹⁸. Cohesin colocalization with the insulator protein CTCF^{8,9} is found at the *Dad1* site 10 kb downstream of *Eα*, whereas at the *Jα49* promoter cohesin associates with its loading protein Nipbl and mediator subunits⁴ more than with CTCF (Supplementary Fig. 4). Interestingly, the major *Tcr* regulatory elements *Eα* and TEA bound copious amounts of cohesin, Nipbl and mediator as well as CTCF (Supplementary Fig. 4).

RNA sequencing (RNA-seq) indicated that *Tcr* constant region (*Cα*) transcripts were considerably more abundant than transcripts from the neighbouring *Dad1* gene and the *Tcrd* constant region (*Cδ*) in

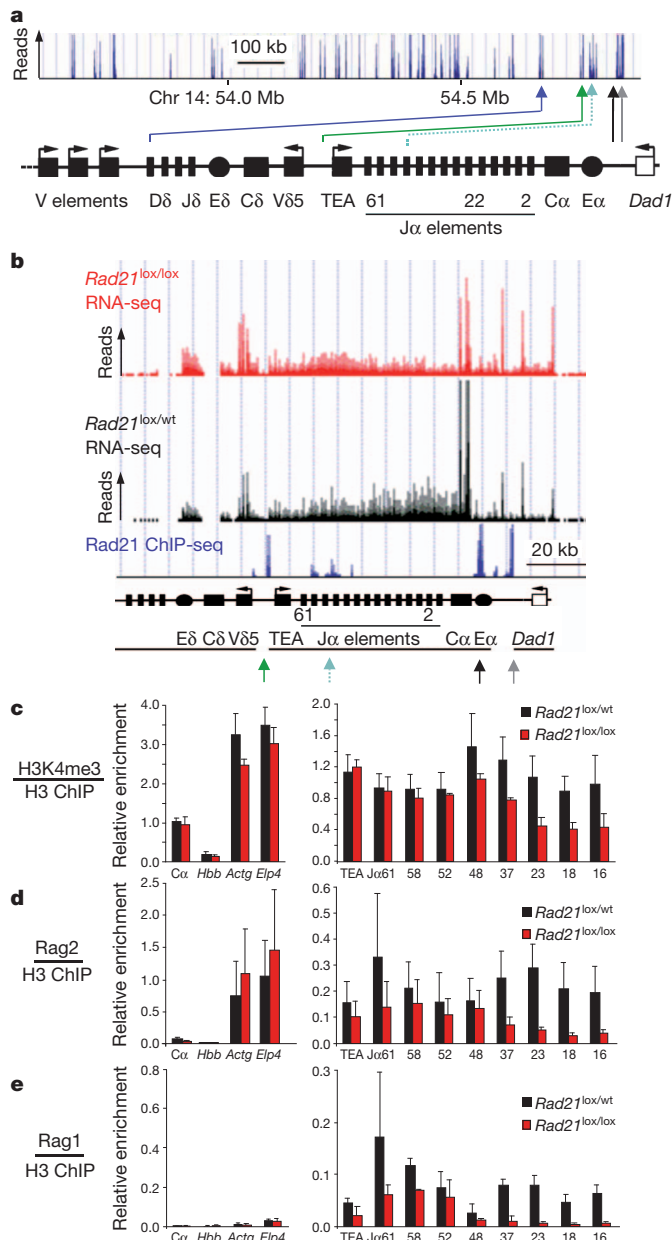


Figure 2 | Cohesin affects *Tcra* transcription and Rag recombinase recruitment. **a**, Rad21 ChIP-seq of the 3' part of the *Tcra* locus in DP thymocytes. Arrowheads highlight cohesin sites at the E α enhancer (black), the *Tcra* locus control region (grey), J α promoters (turquoise), the TEA promoter (green) and between *Tcrd* elements and V gene segments (blue). **b**, RNA-seq of *Tcra* in CD4-Cre *Rad21*^{lox/lox} (red) and control *Rad21*^{lox/wt} (black) DP thymocytes. Rad21 ChIP-seq is in blue. **c**, ChIP of H3K4me3 relative to total H3 in CD4-Cre *Rad21*^{lox/lox} and control DP thymocytes. *Hbb* is a negative and *Actg* and *Elp4* are positive control loci (mean \pm s.e. of two independent experiments). $P = 0.016$ for all J α elements; $P = 0.38$ (not significant) for proximal (J α 61-48) and $P = 0.004$ for distal (J α 37-16) J α elements. **d**, ChIP of Rag2 relative to total H3 as in **c** (mean \pm s.e. of three independent experiments). $P = 0.0001$ for all J α elements; $P = 0.054$ (not significant) for proximal (J α 61-48) and $P = 0.0001$ for distal (J α 37-16) J α elements. **e**, ChIP of Rag1 relative to total H3 as in **c** (mean \pm s.e. of two independent experiments). $P = 0.005$ for all J α elements; $P = 0.17$ (not significant) for proximal (J α 61-48) and $P = 0.003$ for distal (J α 37-16) J α elements.

control DP thymocytes. In cohesin-depleted small DP thymocytes, C δ and *Dad1* transcripts were elevated at the expense of C α transcripts (Fig. 2b) as confirmed by real-time polymerase chain reaction with reverse transcription (RT-PCR) (Supplementary Fig. 5a). Moreover, transcription across the *Tcra* joining elements, J α , was skewed: control DP thymocytes preferentially transcribed distal (3') J α elements, whereas cohesin-depleted DP thymocytes preferentially transcribed proximal (5') J α elements (Fig. 2b).

Transcription of lymphocyte receptor loci facilitates rearrangement²⁷ in part via the trimethylation of histone H3 at lysine 4 (H3K4me3). H3K4me3 recruits Rag2 protein^{14,15,28}, which together with Rag1 forms the recombinase complex¹⁸. ChIP showed reduced H3K4me3 deposition (Fig. 2c) and Rag binding (Fig. 2d, e) at distal J α elements in CD4-Cre *Rad21*^{lox/lox} DP thymocytes. Hence, cohesin deficiency affected *Tcra* transcription, H3K4me3 histone modifications and the recruitment of Rag recombinases.

Primary *Tcra* rearrangements involve proximal (5') J α elements and occur in early CD71⁺ DP thymocytes, whereas secondary rearrangements involve progressively more distal (3') J α elements in non-dividing DP thymocytes¹⁸ (Supplementary Fig. 1b). CD71⁺ DP CD4-Cre *Rad21*^{lox/lox} thymocytes had near-normal Rad21 protein levels (Fig. 1c) and C δ , C α and *Dad1* transcription (Supplementary Fig. 5b), and primary rearrangements of proximal J α elements were present at normal levels (Fig. 3a, b). In contrast, secondary *Tcra* rearrangements were substantially impaired in non-dividing CD4-Cre *Rad21*^{lox/lox} DP thymocytes (Fig. 3a, b), which were depleted of cohesin (Fig. 1c). The usage of the distal J α 22 element, for example,

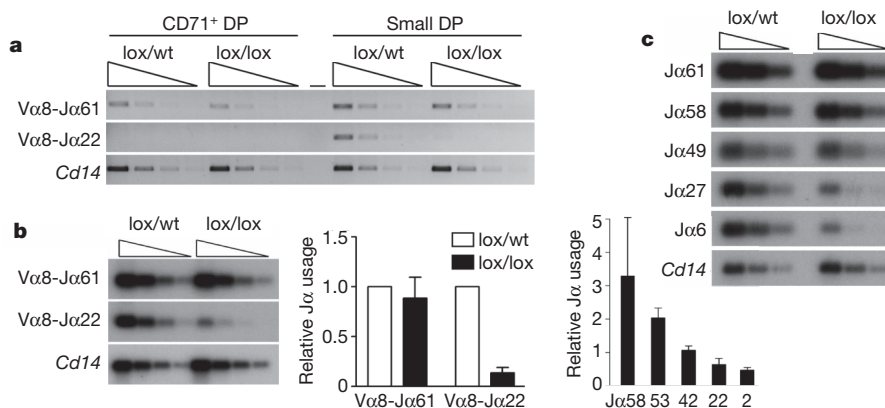


Figure 3 | Cohesin affects *Tcra* rearrangement. **a**, Threefold dilutions of genomic V α 8-J α PCR products from CD71⁺ or CD71⁻ DP thymocytes visualized with ethidium bromide. *Cd14* is a genomic control. **b**, Genomic PCR products from small DP thymocytes visualized by Southern blotting with J α -specific probes (left). Usage of the distal J α 22 element was reduced by 86%

(middle; mean \pm s.e., $n = 3$). Southern blotting of V α 8-J α RT-PCR products from CD4-Cre *Rad21*^{lox/lox} normalized to control DP thymocytes (right; mean \pm s.e., $n = 3$). **c**, Double-strand breaks in threefold serially diluted genomic DNA from CD4-Cre *Rad21*^{lox/lox} and control DP thymocytes detected by ligation-mediated PCR.

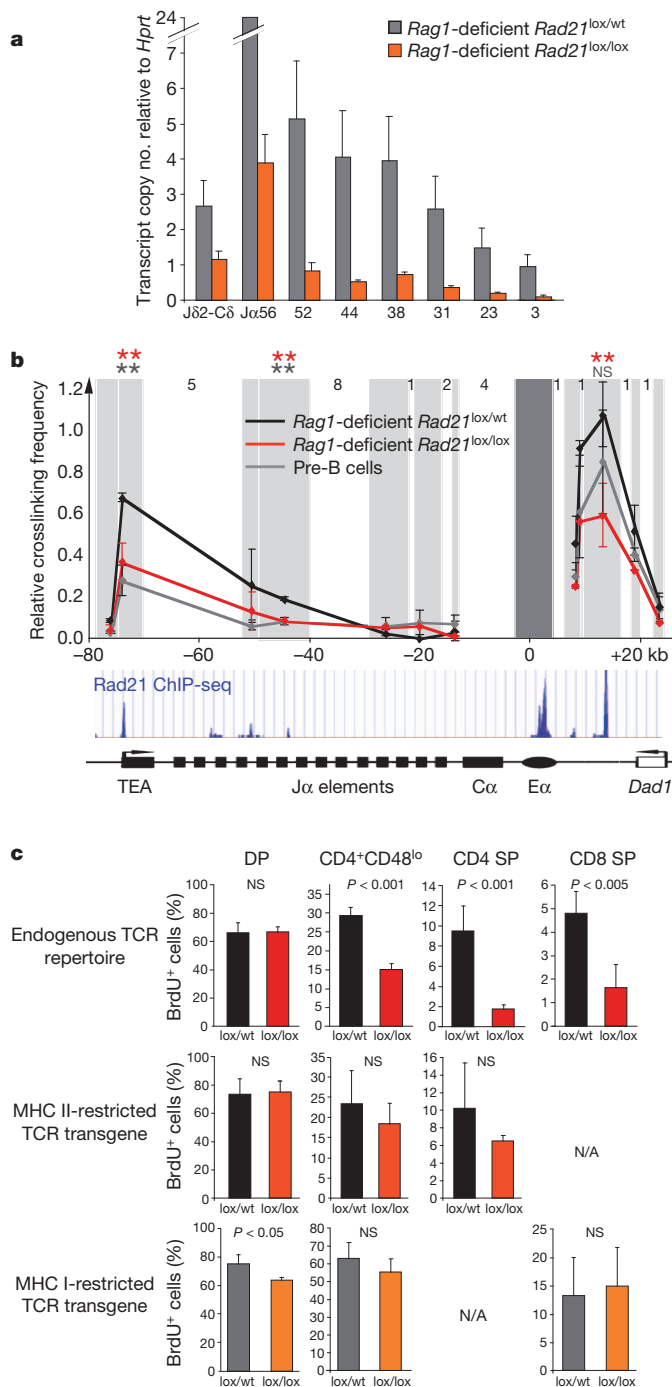


Figure 4 | Cohesin mediates long-range interactions between regulatory elements that control *Tcrα* transcription. **a**, Transcript copy number of the unrearranged *Tcrα* J region in Rag1-deficient CD4-Cre *Rad21*^{lox/lox} and *Rad21*^{lox/wt} DP thymocytes (mean ± s.e., *n* = 3). **b**, 3C analysis of long-range interactions between Eα and Tcrα restriction fragments (shaded) in Rag1-deficient CD4-Cre *Rad21*^{lox/lox} (black; mean ± s.d., *n* = 3), CD4-Cre *Rad21*^{lox/wt} (red; mean ± s.d., *n* = 3) DP thymocytes and pre-B cells (grey; mean ± s.d., *n* = 3). Intervening HindIII fragment numbers and genomic distances are indicated. Asterisks indicate *P* < 0.05 (grey, control thymocytes versus pre-B cells; red, control versus cohesin-depleted thymocytes). NS, not significant. **c**, TCR transgenes rescue the differentiation of cohesin-depleted thymocytes. Top, percentages of BrdU⁺ cells in CD4⁺ CD8^{lo}, CD4 SP and CD8 SP thymocytes. The differentiation of cohesin-deficient thymocytes is rescued by MHC class II-restricted (middle) and MHC class I-restricted TCR transgenes (bottom). N/A, not applicable. *n* = 3–5 per data point ± s.d.

was reduced on average by 86% (Fig. 3b, middle), reflecting a progressive underrepresentation of 3' Jα segments (Fig. 3b, right). This was confirmed by analysis of additional Vα gene families (Supplementary Fig. 6a) and by RT-PCR-based copy number analysis of mature *Tcrα* transcripts (Supplementary Fig. 6b). Therefore, primary rearrangements occurred before the depletion of cohesin, whereas reduced cohesin expression impaired secondary rearrangements in non-dividing DP thymocytes. Normal BrdU labelling kinetics (Fig. 1f) exclude decreased lifespan as an explanation for aberrant *Tcrα* rearrangement²⁹. Consistent with the defective recruitment of Rag proteins to the *Tcrα* locus (Fig. 2d, e), double-strand breaks were reduced in cohesin-deficient thymocytes (Fig. 3c). This identifies Rag cleavage rather than double-strand break repair as the limiting step for *Tcrα* rearrangements in cohesin-deficient thymocytes.

Because *Tcrα* rearrangement changes the positioning of regulatory elements¹⁸, altered *Tcrα* transcription (Fig. 2b) could either be a direct consequence of cohesin depletion, or result indirectly from defective rearrangement. To distinguish between these possibilities we compared intronic *Tcrα* transcript copy numbers in the absence of rearrangement in Rag1-deficient control and CD4-Cre *Rad21*^{lox/lox} DP thymocytes. Cohesin depletion reduced the transcription (Fig. 4a) and H3K4me3 methylation (Supplementary Fig. 7) at Jα independently of *Tcrα* rearrangement.

To explore how cohesin affects *Tcrα* transcription, we analysed long-range interactions between the TEA promoter and Eα, which are separated by approximately 80 kb of genomic DNA and together regulate the transcription of *Tcrα*. In chromosome conformation capture (3C) assays³⁰, Eα interacted strongly with TEA in DP thymocytes (Fig. 4b) but cohesin depletion reduced these interactions to the level found in pre-B cells, where *Tcrα* is not detectably transcribed (Fig. 4b). Hence, the extent of *Tcrα* enhancer–promoter interactions was cell-type specific, correlated with *Tcrα* transcription and was cohesin dependent. A role for cohesin in additional enhancer–promoter interactions during sequential *Tcrα* rearrangements is suggested by Eα contacts with promoters between Jα49 and Jα37 (Fig. 4b), which can drive Jα transcription in the absence of TEA¹⁸, and by cohesin binding to numerous Vα promoters (Fig. 2a). Cohesin depletion also affected Eα interactions with the neighbouring *Dad1* cohesin site, and therefore the topology of the *Tcrα* locus control region²⁴, which has CTCF-dependent transcriptional insulator function^{25,26} (Fig. 4b). As cohesin mediates CTCF-dependent transcriptional insulation^{8,9}, increased *Dad1* expression at the expense of Cα (Fig. 2b) may indicate impaired insulator function.

To test whether aberrant *Tcrα* rearrangement caused inefficient differentiation, we equipped cohesin-deficient thymocytes with transgenes encoding rearranged TCRs. Compared to endogenously rearranged TCRs (Fig. 4c, top), the expression of MHC class II- (Fig. 4c, middle) or MHC class I-restricted TCRs (Fig. 4c, bottom) markedly improved the generation of CD4-Cre *Rad21*^{lox/lox} CD4⁺ CD8^{lo} and SP thymocytes.

In summary, cohesin shapes the chromatin architecture of the *Tcrα* locus by mediating cell-type-specific long-range interactions between enhancer and promoter elements that control transcription, H3K4me3 deposition, Rag recombinase recruitment, and ultimately *Tcrα* rearrangement. These defects compromise thymocyte differentiation by limiting the number and diversity of sequential *Tcrα* rearrangements. Hence, cohesin contributes to cellular differentiation in a well-characterized mammalian system.

METHODS SUMMARY

The conditional *Rad21* allele was generated by inserting *loxP* sites into introns 4 and 6 (Supplementary Fig. 1a). Methods used for RT-PCR and genomic PCR², flow cytometry², 3C analysis², ChIP for cohesin², histone modifications³⁰ and Rag proteins³⁰ have been described. See Methods for other mouse strains, BrdU labelling and detection, ChIP-seq and RNA-seq protocols, ligation-mediated PCR, confocal microscopy, *Tcrα* rearrangement assays and copy number measurements.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2010; accepted 20 June 2011.

Published online 10 August 2011.

1. Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
2. Hadjur, S. *et al.* Cohesins form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus. *Nature* **460**, 410–413 (2009).
3. Degner, S. C. *et al.* CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. *Proc. Natl Acad. Sci. USA*. doi:10.1073/pnas.1019391108 (23 May 2011).
4. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
5. Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor and glue. *Nature Rev. Genet.* **4**, 520–534 (2003).
6. Rollins, R. A., Morcillo, P. & Dorsett, D. Nipped-B, a *Drosophila* homologue of chromosomal adherins participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**, 577–593 (1999).
7. Strachan, T. Cornelia de Lange syndrome and the link between chromosomal function, DNA repair and developmental gene regulation. *Curr. Opin. Genet. Dev.* **15**, 258–264 (2005).
8. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
9. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
10. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
11. Schuldiner, O. *et al.* piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev. Cell* **14**, 227–238 (2008).
12. Pauli, A. *et al.* Cell-type-specific TEV protease cleavage reveals cohesin functions in *Drosophila* neurons. *Dev. Cell* **14**, 239–251 (2008).
13. Pauli, A. *et al.* A direct role for cohesin in gene regulation and ecdysone response in *Drosophila* salivary glands. *Curr. Biol.* **20**, 1787–1798 (2010).
14. Matthews, A. G. *et al.* RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106–1110 (2007).
15. Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R. & Desiderio, S. A plant homeodomain in Rag-2 that binds hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**, 561–571 (2007).
16. Kiehl, P. & von Boehmer, H. Development and selection of T cells: facts and puzzles. *Adv. Immunol.* **58**, 87–209 (1995).
17. Stanhope-Baker, P., Hudson, K. M., Shaffer, A. L., Constantinescu, A. & Schlissel, M. S. Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity *in vitro*. *Cell* **85**, 887–897 (1996).
18. Krangel, M. S. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* **21**, 133–139 (2009).
19. Jhunjhunwala, S., van Zelm, M. C., Peak, M. M. & Murre, C. Chromatin architecture and the generation of antigen receptor diversity. *Cell* **138**, 435–448 (2009).
20. Desiderio, S., Lin, W. C. & Li, Z. The cell cycle and V(D)J recombination. *Curr. Top. Microbiol. Immunol.* **217**, 45–59 (1996).
21. Brekelmans, P. *et al.* Transferrin receptor expression as a marker of immature cycling thymocytes in the mouse. *Cell. Immunol.* **159**, 331–339 (1994).
22. Huesmann, M., Scott, B., Kiehl, P. & von Boehmer, H. Kinetics and efficacy of positive selection in the thymus of normal and T cell receptor transgenic mice. *Cell* **66**, 533–540 (1991).
23. Lee, P. P. *et al.* A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763–774 (2001).
24. Diaz, P., Cado, D. & Winoto, A. A locus control region in the T cell receptor α/δ locus. *Immunity* **1**, 207–217 (1994).
25. Magdinier, F., Yusufzai, T. M. & Felsenfeld, G. Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor α and *Dad1* genes. *J. Biol. Chem.* **279**, 25381–25389 (2004).
26. Zhong, X. P. & Krangel, M. S. Enhancer-blocking activity within the DNase I hypersensitivity site 2 to 6 region between the T cell receptor α and *Dad1* genes. *J. Immunol.* **163**, 295–300 (1999).
27. Abarrategui, I. & Krangel, M. S. Germine transcription: a key regulator of accessibility and recombination. *Adv. Exp. Med. Biol.* **650**, 93–102 (2009).
28. Ji, Y. *et al.* The *in vivo* pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419–431 (2010).
29. Guo, J. *et al.* Regulation of the TCR α repertoire by the survival window of CD4⁺CD8⁺ thymocytes. *Nature Immunol.* **3**, 469–476 (2002).
30. Dekker, J. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nature Methods* **3**, 17–21 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Hadjur, D. Tough, L. Williams, Z. Webster, J. Godwin and H.-Y. Shih for help and advice, L. Game and M. Jones for high-throughput sequencing, A. Giess for sequence alignment, and J. Elliott and P. Hexley for cell sorting. Supported by the Medical Research Council, UK (V.S., T.L., H.M.-B., K.E.B., T.C., A.T., L.A., A.G.F., K.N., M.M.), the European Union FP6 integrated project HEROIC (H.M.), EU and the Marie Curie Research Training Network Chromatin Plasticity (H.M.-B.), the Boehringer Ingelheim Fonds (T.L.), the Wellcome Trust (D.J.A., K.N.) and the National Institutes of Health (B.H., M.S.K., G.T., D.G.S.). D.G.S. is an investigator of the Howard Hughes Medical Institute.

Author Contributions V.S. and M.M. conceived the study with critical input from D.G.S., L.A., A.G.F., M.S.K. and K.N., V.S., B.H., K.T.-K., T.L., H.M.-B., K.E.B., G.T., K.H. and M.M. conducted experiments, K.T.-K., D.J.A., K.N., G.T. and D.G.S. designed and generated novel materials, T.C., A.T. and H.M. analysed data, V.S. and M.M. wrote the paper and all authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.M. (matthias.merkenschlager@csc.mrc.ac.uk), K.M. (kim.nasmyth@bioch.ox.ac.uk) or M.S.K. (krang001@mc.duke.edu).

METHODS

Mice. *Rag1*-deficient mice³¹, OT-I as well as AND TCR transgenic mice^{32,33} have been described. To generate a conditional *Rad21* allele, *loxP* sites were inserted into introns 4 and 6 to allow Cre-mediated excision of exons 5 and 6. Correctly targeted clones were detected by Southern blotting of BmtI and PacI digested genomic DNA to confirm the presence of 5' and 3' *loxP* sites, respectively. A *puroAtk* cassette flanked by *FRT* sites was excised to generate the conditional allele (Supplementary Fig. 1a). For genomic deletion analysis DNA was extracted and 50 ng were used per real-time PCR reaction. Primers were designed to detect the intact and not the deleted *Rad21*^{lox} allele. Data were normalized to the geometric mean of two genomic DNA (gDNA) sites on chromosome 2 and 3, respectively.

Rad21^{lox}, forward, CATGGTTGGCAGATGAGCAAC; *Rad21*^{lox}, reverse, CTACTTTCCCGTAGCAACTG; gDNA chr 2, forward, CCCGGGCTAAT TCTCTATGTC; gDNA chr 2, reverse, GCTGTAAAGTCAGTCGCTCGTG; gDNA chr 3, forward, TGTGCCAGCATCTTTGCC; gDNA chr 3, reverse, GCGTGTGGAATTAAGGCC.

Mice were on a mixed C57BL/6 and 129 background and were matched for age and *Tcr* haplotypes in all experiments involving *Tcr* expression or rearrangement. *Tcr* haplotypes were determined by genomic PCR.

Primer 1, GAGGAAAAATGGCCCGTAG; primer R, TTCAGTAGTCCCT CTCCACG. PCR amplification 94 °C for 5 min, 5 cycles of 94 °C for 30 s, 68 °C for 30 s, 72 °C for 1 min, 5 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 1 min, 5 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min, 20 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, 10 min extension at 72 °C, HindIII digestion yields a 300 bp band for 129 and a 150 bp band for C57BL/6.

Rag1-deficient thymocytes were induced to progress to the CD4⁺ CD8⁺ DP stage by injecting 4–6 week-old *Rag1*-deficient mice once with 100 µg anti-CD3 (145-2C11) intraperitoneally. Thymocytes were harvested 10 days later. For BrdU labelling experiments, mice were given a single intraperitoneal injection of 2 mg BrdU (pulse-chase) or 0.6 mg BrdU every 12 h (continuous labelling). Mouse work was done under project licence PPL70/6845 issued by the Home Office, UK.

J α segment usage was assessed by RT-PCR using V α 8 and C α primers, following which Southern blots of PCR products were hybridized with radiolabelled J α -specific and C α -specific probes³⁴.

Ligation-mediated PCR. This was done as described³⁵ using J α -specific primers and probes (5' to 3'). J α 61, primer (internal), AGTCCCTTCTCAGGCAAGAT GG; probe, TGAGGAACACGGAGTATCTC. J α 58, primer, ATGGCTTTGGAC CATGGATG; probe, TCTGAAACTCGACAGTGGA. J α 49, primer, AGGGAA AGTGACACCAGG; probe, AGGTGTAGATTCTAGTCTG; J α 27, primer, ATGGCAGATAGAATGGAGCGG; probe, TACCTCCACCTGTCTTCTCA. J α 6, primer, ATCAGACCAGACTGTCTGCC; probe, GACCAATGGCAAAG GGAGGT.

Confocal analysis. Mitotic figures were stained for α -tubulin (DM1A, Sigma) and DAPI as described³⁶.

Real-time PCR. Real-time PCR was carried out in a 20 µl reaction volume using iQ SYBR Green Supermix (Bio-Rad) and 0.5 µM primers. Amplification was 95 °C for 3 min, then 40 cycles of 94 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s followed by plate read.

RT-PCR and transcript copy number analysis. Total RNA was extracted using RNeasy (Tel-Test) and 1 µg was treated with 1 µl Turbo DNase (Ambion). Following DNase inactivation the RNA was reverse transcribed using 200 U Superscript III (Invitrogen) and 0.5 µg oligo(dT) or 3 µg random primers in a 20 µl reaction volume. 1/80 was used per real-time PCR reaction.

To estimate the number of transcripts, standard curves were generated for each pair of primers. The PCR amplicons generated by individual primer pairs were separated by agarose gel electrophoresis, gel purified and their concentration was determined using Quant-iT PicoGreen (Invitrogen). The molar weight of the amplicons was calculated based on their sequence using OligoCalc³⁷, which allowed the enumeration of copies in each sample. Twofold serial dilutions from 2,500,000 to 1,220 copies (spliced transcripts) and 2,000,000 to 976 copies (intronic transcripts) were analysed by real-time PCR. The standard curves were compiled by plotting the real-time PCR C_t values against seeded copy numbers and subsequently fitting a logarithmic function. Only primers that generated a log fit $R^2 > 0.99$ were used. These standard curves were subsequently used to extrapolate the number of transcripts in the biological samples based on the C_t values obtained after real-time RT-PCR. For the analysis of spliced transcripts the forward primer was specific to the respective J α segment and the reverse primer to C α .

eJ α 56, forward, CTGGAGGCAATAATAAGCTGACT; eJ α 52, forward, AACACTGGA AAGCTCACGTTTGG; eJ α 44, forward, GGCAGTGGTGGAAA ACTCACT; eJ α 38, forward, AGCTGATTTGGGGCTGGGGA; eJ α 31, forward, GGACGCAGCTGGTGGTGAAGC; eJ α 23, forward, GCTTATCTTTGGACAG GGAACCA; eJ α 18, forward, AACTTTTGAATCGGAACCAGGG; eJ α 9, forward,

ACAAACTTACCTTCGGGACAGG; TCR α -C, reverse, TCCATAGCTTTTCATGT CCAGC.

Germine transcripts were analysed using primer pairs specific to introns adjacent to the J α segments included in the spliced transcripts analysis. No amplification was detected without reverse transcriptase.

iJ α 56, forward, CCCTTGAACCCGTGATATGC; iJ α 56, reverse, CAGCCATT GTTTGGATTGGA; iJ α 52, forward, CTGTGGGTCTGTGCTTTCTG; iJ α 52, reverse, AGGGAATGGTGGGTACACG; iJ α 44, forward, GAAGAGGCCTC ATGGAGGAA; iJ α 44, reverse, TCTCACAGACCGAAGGGACA; iJ α 38, forward, AGGACTTGGGGTCATCTCCA; iJ α 38, reverse, GGCTCCTCAGGACAGAC CAG; iJ α 31, forward, GCAGCCTGCCAGCTATCTTT; iJ α 31, reverse, CAAA GCGACCAACCAACAA; iJ α 23, forward, AGAGGAGGCCGAAAGTCTCC; iJ α 23, reverse, TGCACAAATCCAGGCCTATG; iJ α 3, forward, TCTGGACCC TTGGCAATCAT; iJ α 3, reverse, CCCTGCCCTGGTCTACTGTG; iJ α 2-C δ , forward, GGGTTTGTAGTTGGCTTTGG; iJ α 2-C δ , reverse, GGCATGTATTTT GCGGTTGA.

ChIP-seq. Single-read high-throughput sequencing libraries were prepared from 10 ng of ChIP DNA and sequenced according to the manufacturer's protocols (Illumina Genome Analyser II). Reads of 38 bases were aligned to the *Mus musculus* mm9 genome assembly using ELAND (Illumina) and wig files generated with FindPeaks³⁸ were visualized with the UCSC Genome Browser (<http://www.genome.ucsc.edu/>).

RNA sequencing. 10 µg total RNA were depleted of ribosomal RNA using RiboMinus (Invitrogen). Depletion was verified using an Agilent 2100 Bioanalyser and the retention of mRNA by real-time RT-PCR. The samples were treated with DNaseI, and the RNA was fragmented for 2 min at 94 °C in RNA fragmentation buffer (final concentration: 40 mM Tris acetate, pH 8.2, 100 mM potassium acetate and 30 mM magnesium acetate). The reaction was stopped by placing the samples on ice and the fragmented RNA was column purified (RNAeasy Minelute Kit, Qiagen), eluted in 10 µl, and used as template for cDNA synthesis using 5 µg random hexamers in a total volume of 20 µl. Second-strand synthesis was performed by adding 91.8 µl water, 30 µl 5* Second strand buffer (Invitrogen), 3 µl 10 mM dNTPs, 4 µl *E. coli* polymerase I (10 U µl⁻¹), 1 µl *E. coli* DNA ligase (10 U µl⁻¹) and 0.2 µl RNase H (10 U µl⁻¹), followed by 2 h at 16 °C. 1 µl T4 DNA Polymerase (10 U µl⁻¹) was added followed by an additional 10 min at 16 °C. The double-stranded cDNA was purified using the Minelute Reaction Cleanup Kit (Qiagen), eluting twice in 10 µl. Following quantification, 20 ng double-stranded cDNA was used for the preparation of a single-read high-throughput sequencing library according to Illumina protocols.

Rearrangement frequency analysis by semi-quantitative PCR. Threefold serial dilutions of 100 ng genomic DNA were used for PCR amplification with forward primers specific to V α 5 or V α 7 and reverse primers specific to J α 56, J α 38 or J α 27, respectively³⁹ (this reference refers to V α 5 as TRAV3-4 and V α 7 as TRADV15-1). Primers for the Calreticulin locus were used as a loading control. PCR products were separated by agarose gel electrophoresis and stained with ethidium bromide.

V α 7, forward, GCAGGATCTAATGTGGCCAGAAAGTGATTCA; V α 5, forward, AGGTGATCACAGAGGCATCCT; J α 56, reverse, ACGTACCTGGTA TAACACTCAGAAC; J α 38, reverse, CAAAGACGACTTTGTTGG; J α 27, reverse, TTAAGAGCCCAAGCAGATGCATAAG; calreticulin, forward, TCAT GAGTTCCCCACATCTTTG; calreticulin, reverse, CTGCCCTATCCTGAGTC TGACA.

3C. Cells were fixed in 10% FCS, 1% formaldehyde for 10 min at room temperature (22 °C) and fixation was stopped with glycine (0.125 M). 10⁷ cells per sample were lysed in 10 mM Tris, pH 8, 10 mM NaCl, 5 mM MgCl₂, 0.2% NP-40 for 30 min on ice. The nuclei were pelleted and re-suspended in 0.5 ml 1.2 \times digestion buffer (NEB2, New England Biolabs) and permeabilized with SDS (0.5% final concentration) for 1 h at 37 °C, shaking at 800 r.p.m. and 3.3% Triton X-100 were added for an additional 1 h at 37 °C. 2,000 U HindIII (New England Biolabs) were added before incubation over night at (37 °C, 800 r.p.m.) and inactivated with SDS (1.5%, 65 °C, 30 min). The reaction was diluted in 6.2 ml 1.1 \times T4 ligase buffer (New England Biolabs) and incubated at 37 °C for 1 h after addition of 1% Triton X-100. 800 U T4 DNA ligase (New England Biolabs) was added for 4 h at 16 °C, crosslinking was reversed by 300 µg proteinase K (65 °C, 16 h). 300 µg RNase A was added for 1 h at 37 °C. DNA was isolated by phenol/chloroform extraction and ethanol precipitation, quantified using Quant-iT PicoGreen (Invitrogen) and 200 ng DNA were used per TaqMan PCR reaction (QuantiFast, Qiagen). Data were normalized to the cross-linking frequency between the anchor and the neighbouring HindIII fragment. HindIII digestion efficiency was calculated as described² (Supplementary Fig. 8a). The efficiency and linearity of 3C primers was tested on templates obtained by HindIII digestion and religation of genomic PCR products spanning the HindIII restriction sites. The products of each primer pair were gel purified and quantified to calculate copy numbers. Standard curves were constructed by tenfold serial dilution (2 \times 10⁶ to 20 copies). 3C primers were

selected for >90% efficiency and efficiency and a logarithmic coefficient of determination (R^2) >0.99 (Supplementary Fig. 8b, c).

Distance from anchor, -75 kb; 3C primer, CTGCCCCAAAAGAGAGCTTGG; reverse primer, GCATGTGAATGAGAGAGGATGG. Distance from anchor, -73 kb (TEA); 3C primer, GCGGATTTGGGCAAGAATAA; reverse primer, CACAGGCTGCTCCAGAGATG. Distance from anchor, -50 kb; 3C primer, TCTCGAAGGTCCACAAAGCA; reverse primer, GGACACGCTTCCTTGGCTAC. Distance from anchor, -44 kb; 3C primer, AGTCGTGTGGCTGAACCTTTC; reverse primer, TGGTGTCTTTATGGGGTTAGGA. Distance from anchor, -26 kb; 3C primer, TTGGCCAGGGGACCATATTA; reverse primer, CTTGCTTC CCGTGATGTCTG. Distance from anchor, -19 kb; 3C primer, CACAGGAGAAATGAGGCTCAGA; reverse primer, CTGACCAGGCCTTCAGATAA. Distance from anchor, -13 kb; 3C primer, TGTCTCTGCCACAAGTGCT; reverse primer, TAGGAGGACGAGAGGCTGGA. Distance from anchor, 0 kb (anchor); 3C primer, AGCATGGGAGACAGACGTAATG; reverse primer, ATTCCGGTCTCCGTGTGAAT. Normalization fragment, 3C primer, ACCTAGTGACGCCCTG GTGA; reverse primer, TTGGCAGGAGTCGGTTCTTT. Distance from anchor, 8 kb; 3C primer, AATGCGGGCATCTCAGACTT; reverse primer, GAGACTCCCAGGCCTTACCC. Distance from anchor, 9 kb; 3C primer, GCACTTGTCTTA TACTGCCCAAGA; reverse primer, ATTGTCTGGAACGGCACCA. Distance from anchor, 13 kb; 3C primer, ATTGTCTGGAACGGCACCA; reverse primer, GCACTTGTCTTATACTGCCCAAGA. Distance from anchor, 19 kb; 3C primer, GGGGGAAAACAACACATCAGA; reverse primer, GGGTCATCATTCATCCC

AGTC. Distance from anchor, 23 kb; 3C primer, TTTAAAAGGCCAGACTGCAACA; reverse primer, CTCCCACATCCATTTCACCA. TaqMan probe, 3C primer, [FAM]-CCTGCCTGCCTGAG GACTGC-[BHQ1].

31. Spanopoulou, E. *et al.* Functional immunoglobulin transgenes guide ordered B-cell differentiation in Rag-1-deficient mice. *Genes Dev.* **8**, 1030–1042 (1994).
32. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
33. Kaye, J. *et al.* Selective development of CD4⁺ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor. *Nature* **341**, 746–749 (1989).
34. Abarrategui, I. & Krangel, M. S. Regulation of T cell receptor- α gene recombination by transcription. *Nature Immunol.* **7**, 1109–1115 (2006).
35. Jackson, A., Kondilis, H. D., Khor, B., Sleckman, B. P. & Krangel, M. S. Regulation of T cell receptor β allelic exclusion at a level beyond accessibility. *Nature Immunol.* **6**, 189–197 (2005).
36. Toyoda, Y. & Yanagida, M. Coordinated requirements of human topo II and cohesin for metaphase centromere alignment under Mad2-dependent spindle checkpoint surveillance. *Mol. Biol. Cell* **17**, 2287–2302 (2006).
37. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).
38. Fejes, A. P. *et al.* FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**, 1729–1730 (2008).
39. Lee, Y. N. *et al.* Differential utilization of T cell receptor TCR α /TCR δ locus variable region gene segments is mediated by accessibility. *Proc. Natl Acad. Sci. USA* **106**, 17487–17492 (2009).

A young source for the Hawaiian plume

Alexander V. Sobolev^{1,2,3}, Albrecht W. Hofmann^{2,4}, Klaus Peter Jochum², Dmitry V. Kuzmin^{2,5} & Brigitte Stoll²

Recycling of oceanic crust through subduction, mantle upwelling, and remelting in mantle plumes is a widely accepted mechanism to explain ocean island volcanism¹. The timescale of this recycling is important to our understanding of mantle circulation rates. Correlations of uraniumogenic lead isotopes in lavas from ocean islands such as Hawaii or Iceland, when interpreted as model isochrons, have yielded source differentiation ages between 1 and 2.5 billion years (Gyr)^{2–5}. However, if such correlations are produced by mixing of unrelated mantle components⁶ they will have no direct age significance. Re–Os decay model ages take into account the mixing of sources with different histories^{7,8}, but they depend on the assumed initial Re/Os ratio of the subducted crust, which is poorly constrained because of the high mobility of rhenium during subduction⁹. Here we report the first data on ⁸⁷Sr/⁸⁶Sr ratios for 138 melt inclusions in olivine phenocrysts from lavas of Mauna Loa shield volcano, Hawaii, indicating enormous mantle source heterogeneity. We show that highly radiogenic strontium in severely rubidium-depleted melt inclusions matches the isotopic composition of 200–650-Myr-old sea water. We infer that such sea water must have contaminated the Mauna Loa source rock, before subduction, imparting a unique ‘time stamp’ on this source. Small amounts of seawater-derived strontium in plume sources may be common but can be identified clearly only in ultra-depleted melts originating from generally highly (incompatible-element) depleted source components. The presence of 200–650-Myr-old oceanic crust in the source of Hawaiian lavas implies a timescale of general mantle circulation with an average rate of about 2 (±1) cm yr^{–1}, much faster than previously thought.

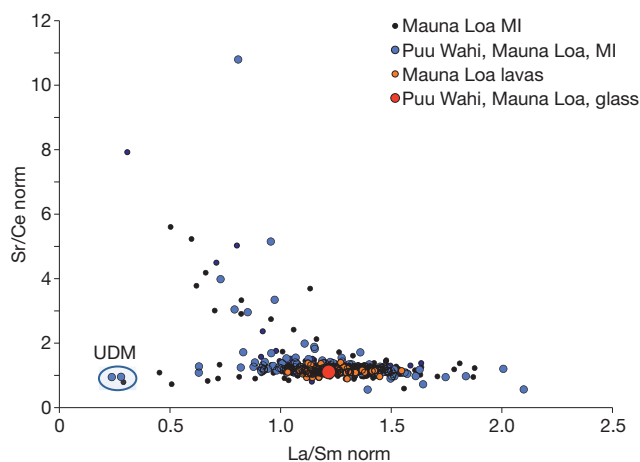


Figure 1 | Compositions of lavas and melt inclusions in olivine phenocrysts from recent (younger than 50 kyr) eruptions of Mauna Loa volcano, Hawaii. Data for Puu Wahi lava and melt inclusions (MI) are from this study; other melt inclusions are from ref. 13 and unpublished data of A.V.S. UDM inclusions are outlined. Compositions of Mauna Loa lavas are from the GEOROC database (<http://georoc.mpch-mainz.gwdg.de/georoc/>). All ratios are normalized to primitive mantle³¹.

Melt inclusions in highly magnesian olivine are commonly used as proxies for parental melts¹⁰. In addition, melt inclusions provide unique information on the isotope heterogeneity of the mantle sources of ocean island basalts^{11,12}. Compositions of melt inclusions in olivine (Supplementary Figs 1 and 2) from lavas of the largest Hawaiian shield volcano, Mauna Loa, vary considerably¹³ (Fig. 1). Both La/Sm and Sr/Ce ratios of melt inclusions have very large (tenfold) variations, and this heterogeneity is much greater than those displayed by Mauna Loa bulk lava compositions. About 2–3% of melt inclusions are strongly depleted in the highly incompatible elements Th, U, Ba, Rb, K, Sr, Nb, Ta, Cl, B, Pb, S, Zr, Hf and the light rare-earth elements (Figs 1 and 2 and Supplementary Table 2). Here we will call these inclusions ‘ultra-depleted melts’ (UDMs). In contrast with typical Mauna Loa melts, UDMs are also slightly depleted in the moderately incompatible elements Ti, Nd, Sm and Al and are enriched in SiO₂, but are similar in contents of heavy rare-earth elements, Y, Sc and Ca (Fig. 2, Supplementary Fig. 3 and Supplementary Table 2). Strontium is significantly depleted relative to Pr, Nd and Sm. UDM inclusions have been reported in association with ‘normal’ inclusions in the same olivine grains¹³. In such an association, inclusions of mixed composition are commonly present. Mauna Loa UDMs have been interpreted¹³ as instantaneous melts produced by near-fractional melting of Hawaiian source by a process that is similar to that described for mid-ocean-ridge UDM¹⁴. The isotope data presented here disprove this interpretation, showing that these UDMs also have significantly different source compositions from those of other parental melts.

Strontium and lead isotopic compositions of melt inclusions in olivine phenocrysts from a single Mauna Loa lava (see Supplementary Figs 1 and 2 for inclusion images) were measured by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS; see

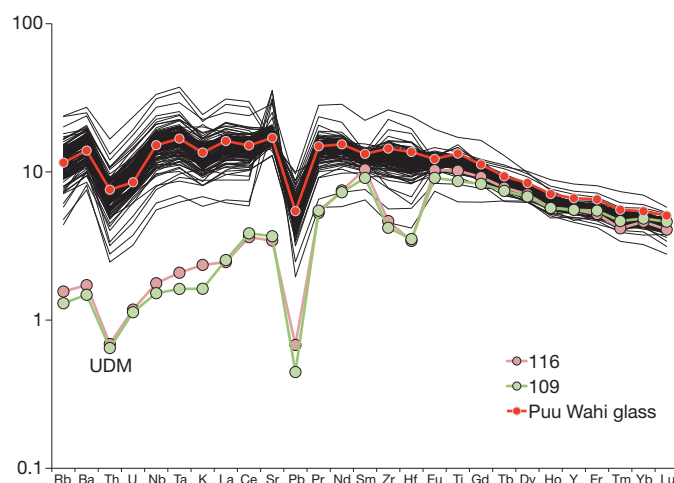


Figure 2 | Primitive mantle³¹ normalized concentrations of incompatible elements in melt inclusions in euhedral olivine crystals from a single sample (K97-15b) of Puu Wahi scoria cone, Mauna Loa, Hawaii. Ultra-depleted melt inclusions are labelled UDM.

¹ISTerre, University Joseph Fourier, Grenoble 1 and CNRS, BP 53, 38041 Grenoble, France. ²Max Planck Institute for Chemistry, Postfach 3060, 55020 Mainz, Germany. ³V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, Moscow 119991, Russia. ⁴Lamont Doherty Earth Observatory, Palisades, New York 10964, USA. ⁵V. S. Sobolev Institute of Geology and Mineralogy, Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090, Russia.

Methods for sample description and analytical details). In total, 138 inclusions were analysed for Sr isotopes, and 106 inclusions for Pb isotopes. Both Sr and Pb isotopic compositions of melt inclusions vary markedly, much more than in Hawaiian lavas overall. But their averages are nevertheless similar in compositions to those of matrix glass and of typical recent Mauna Loa lavas (Fig. 3). The range of variation is markedly higher for melt inclusions trapped in the most magnesian olivines (Supplementary Fig. 4). This is consistent with the interpretation that melt inclusions in early-formed olivine phenocrysts yield information on the compositions of unmixed parental melts, whereas lavas and inclusions in more evolved olivines are mixtures of these melts^{10,13}. Clear evidence of such mixing is presented by the most radiogenic melt inclusions with $^{87}\text{Sr}/^{86}\text{Sr} > 0.7060$. These inclusions consistently fit mixing trajectories for almost all incompatible elements and Sr isotopes, thus strengthening confidence in the reliability of the isotope data (Supplementary Fig. 5). UDMs represent the endmember of such mixing combining highly radiogenic Sr ($^{87}\text{Sr}/^{86}\text{Sr} = 0.7081 \pm 0.0006$, 2σ), much higher than the value of typical Mauna Loa melts ($^{87}\text{Sr}/^{86}\text{Sr} \approx 0.7038$), with Rb/Sr ratios (0.010–0.013) that are unusually low for Mauna Loa (Fig. 3a). Lead isotope compositions of UDMs are also significantly more radiogenic than those of typical Mauna Loa melts (Fig. 3b; note that the lowest $^{207}\text{Pb}/^{206}\text{Pb}$ ratios correspond to the most radiogenic Pb). These data clearly suggest that the UDMs could not have been produced from any typical Mauna Loa source by any melting process. Instead, the source rock must contain a component with highly radiogenic Sr and Pb, but must at the same time be strongly depleted in incompatible elements. Alternatively, this component might have been introduced by some type of contamination of the primary melt. Next we discuss possible sources.

Before presenting our preferred interpretation, namely source contamination by ancient sea water before subduction, we discuss alternative explanations. In general, the highly radiogenic Sr of Mauna Loa UDMs might be explained by one of the following processes, all of which have significant drawbacks.

First, direct input of recycled continental crust could in principle explain highly radiogenic Sr, but this is completely inconsistent with the extremely low concentrations of Rb, Ba, K, U and Pb, as well as their high Ce/Pb (50–80) and Nb/U (41–46) ratios¹⁵.

Second, continental Sr might be transferred to the mantle source during subduction more indirectly by the percolation of crustal fluids without necessarily causing trace-element enrichment, if the mantle lithology is a depleted harzburgite containing no clinopyroxene. This process requires that, in addition to Sr, all other incompatible elements including Sc, Ca and Al in UDMs come from a clinopyroxene-free harzburgite source. However, this contradicts the relatively high contents of moderately incompatible elements (Ti, Al, Ca and rare-earth elements heavier than Sm) in UDMs that are similar to those of typical Mauna Loa melts (Fig. 2 and Supplementary Fig. 3), because harzburgites are typically depleted in all these elements. Additional observations that are very difficult to reconcile with any continental source, even in a markedly 'fractionated' form, are as follows: the very low Pb concentrations and high Ce/Pb ratios, as well as high Nb/U and Nb/Th ratios, are the opposite of what would be expected from a metasomatic (fluid) transfer of a sedimentary signature from a subducted slab or from recycled sediments (see Supplementary Information); and other ratios of trace elements of similar incompatibility (Ba, Rb, Th, U and Nb) are similar (though not identical) to those of oceanic basalts, in particular Hawaiian basalts, but quite different from those of continental materials.

Third, the radiogenic Sr might be derived directly or indirectly from modern sea water. Producing UDM by contamination with typical Mauna Loa melt is highly unlikely because such a contamination process cannot result in tenfold lower concentrations of most incompatible elements. Seawater assimilation would be somewhat more plausible if it affected the original, highly depleted melt (rather than normal Mauna Loa melt). Perhaps the most efficient mechanism for this would be the assimilation of recent aragonitic carbonate containing a high concentration of radiogenic Sr and depleted in all incompatible elements¹⁶. This process could add radiogenic Sr without significantly changing most of the incompatible-element budget of the melt. However, it would also increase Sr and Ca concentrations considerably. For example, a typical Mauna Loa melt with $^{87}\text{Sr}/^{86}\text{Sr} = 0.7040$ would require the assimilation of about 5% high-Sr carbonate to match the isotopic composition of UDM. This would result in a sixfold increase in Sr concentration of UDM if the carbonate contaminant contained 1,000 p.p.m. Sr, with $^{87}\text{Sr}/^{86}\text{Sr} = 0.7090$. This implies an unreasonably

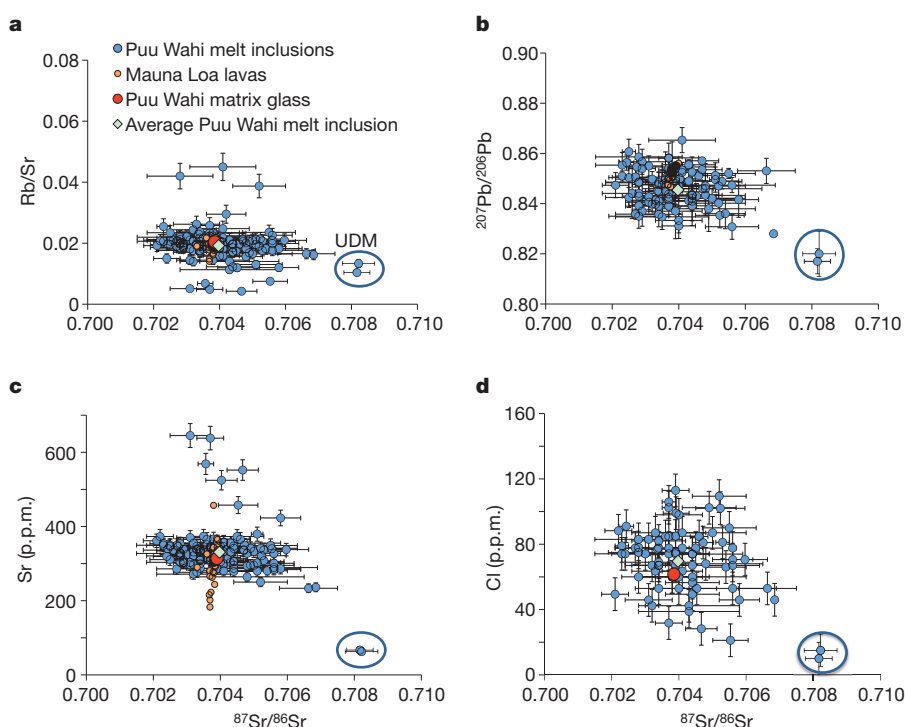


Figure 3 | $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in melt inclusions and matrix glass in euhedral olivine crystals of sample (K97-15b), Puu Wahi scoria cone, Mauna Loa, Hawaii. **a–d**, $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are plotted against Rb/Sr (**a**), $^{207}\text{Pb}/^{206}\text{Pb}$ (**b**), Sr (**c**) and Cl (**d**). Compositions of bulk Mauna Loa lavas from the GEOROC database are shown (in orange) for comparison. The compositions of average UDM inclusions are outlined. Error bars indicate s.e.m.

large negative Sr anomaly (half the primitive mantle value of Sr) in the initial, uncontaminated melt. Also, this process must increase the CaO content of the UDMs by at least 1.5 wt%, which is not observed (see Supplementary Fig. 3). Alternatively, seawater Sr might enter the melt by assimilation of highly altered ocean crust (including sediment). This could explain highly radiogenic Sr (Fig. 4) and low concentrations of incompatible elements, as well as more radiogenic Pb¹⁷. However, this should also incorporate other seawater-specific components, particularly B, Cl and K, because these elements are strongly enriched in altered crust¹⁸. But this is inconsistent with the extremely low concentrations of Cl (10–15 p.p.m.), B (0.3–0.4 p.p.m.) and K₂O (0.05–0.07 wt%) in the UDMs (Fig. 3d and Supplementary Table 2). Even if one assumes that these volatile elements were extracted from the assimilated crustal rocks by previous high-temperature metamorphism, this process still seems unlikely. To raise the originally low $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of the melt to the highly radiogenic value observed, the amount of such assimilated material would have to be large. Such extensive assimilation would also affect concentrations of Ca, the heavy rare-earth elements, Y and Sc, all of which are relatively high and very similar in UDMs and typical Mauna Loa lavas (Fig. 2, Supplementary Fig. 3 and Supplementary Table 2). Alumina is slightly depleted in UDMs in comparison with Mauna Loa melts (Supplementary Fig. 3), and this also precludes the assimilation of any plagioclase-saturated, high-Al rocks, typical of oceanic crust. Similarly, we can rule out the assimilation of seawater-altered Hawaiian lavas, because of their high incompatible-element enrichment.

Reaction with altered harzburgites and their partial melting may also deliver highly radiogenic modern seawater Sr to depleted Mauna Loa

melt. But this process will also affect concentrations of Ni, Mn and Fe, which is not the case (Supplementary Fig. 8).

An additional argument against contamination by the present-day lithosphere beneath Hawaii is derived from the fact that this lithosphere is slightly more than 90 Myr old. At the time that it was created, sea water had a $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.7074 or less (ref. 19), just below the lower limit of UDM Sr (Fig. 4). Because seawater Sr is introduced into the altered crust almost entirely during the first 3 Myr after crust formation²⁰, it is highly unlikely that Sr contamination derived from 90-Myr-old lithosphere can account for the observed composition ($^{87}\text{Sr}/^{86}\text{Sr} \geq 0.7075$) of the UDMs.

Because none of the above processes provides a satisfactory explanation for the presence of highly radiogenic Sr and Pb in the UDM, we suggest that the radiogenic Sr was introduced through direct or indirect contamination by ancient sea water before subduction and recycling. In this case, other indicators of sea water or seawater alteration, K, B and Cl, should easily be removed from the system by mobilization during subduction²¹. In addition, Ca, Al, rare-earth elements, Sc and Y in the melt would be constrained by the mineralogy of the mantle residue, specifically by the presence of garnet and high-Ca pyroxene. However, Sr in ancient sea water was, during most of Earth's history, much less radiogenic than Sr in the UDMs, except for one specific time window, namely the period between 200 and 550 Myr ago^{19,22} (Fig. 4). According to other estimates²³ this period may be extended at maximum to 650 Myr. Therefore, if the effect is indeed caused by recycled seawater Sr, this sea water is unlikely to be older than 650 Myr old.

We propose that Mauna Loa UDMs reveal one of the recycled end-member compositions similar to ultra-depleted mid-ocean-ridge basalt (MORB)¹⁴ or gabbros²⁴ with Sr extensively exchanged with, and dominated by, seawater Sr because of its initially low content (see model in Methods). This explanation is also consistent with the strong negative Zr anomaly in Mauna Loa UDMs, which is a common feature of ultra-depleted MORB as well¹⁴. Highly depleted gabbro-norites crystallized from remelted material of this kind possessing $^{87}\text{Sr}/^{86}\text{Sr}$ up to ~0.708 have been reported in oceanic crust exposed in the Oman ophiolite²⁵. The Sr isotopes of the other recycled rocks (typical gabbros or normal MORBs) indicated by the compositions of melt inclusions (Fig. 3) were less affected by seawater alteration because of their higher Sr contents and/or different positions in the crust¹⁸. The Sr-rich component of Mauna Loa lavas is particularly unradiogenic ($^{87}\text{Sr}/^{86}\text{Sr} < 0.7030$; see Fig. 3c), in agreement with its origin from recycled oceanic plagioclase-cumulate gabbros¹³. Alternatively, seawater Sr might be introduced into the subduction 'package' by means of serpentinites or carbonate veins that are ubiquitous in the oceanic crust. Particularly when precipitated at low temperatures, they preserve the Sr isotopic composition of sea water¹⁶. We further suggest that all these recycled materials have been processed during subduction to remove most of the additional Cl, B, Rb and K introduced by sea water.

Previous estimates of recycling times in mantle plumes have mostly relied on Pb isotopes. If an initially homogeneous source is differentiated (for example, by melting) into subsystems with variable U/Pb ratios, these subsystems evolve to a linear isochron in $^{207}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ space. Such linearly correlated Pb isotopes are widespread in oceanic basalts and have frequently been interpreted as isochrons corresponding to recycling ages ranging from 1.0 to 2.5 Gyr (ref. 2). However, they might also simply represent mixing lines between unrelated reservoirs⁶. It has been pointed out²⁶ that the isochron interpretation fails when Th/U ratios inferred from the slope of $^{208}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ correlations are inconsistent with observed Th/U ratios in the rocks, and it was concluded that the Pb isotope correlations for Mauna Kea must represent mixing lines rather than isochrons.

All Mauna Loa olivine phenocrysts possess Ni excess and Mn deficiencies, which suggests the involvement of melts from olivine-free reaction pyroxenite and peridotite sources in proportions of about

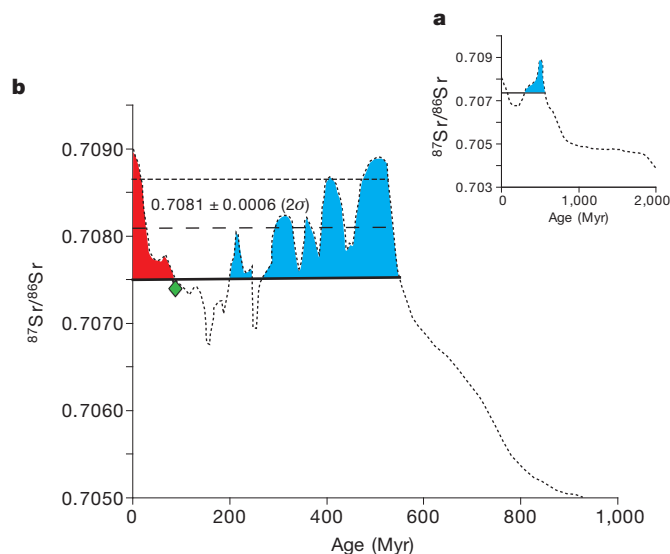


Figure 4 | $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of the two most radiogenic Mauna Loa melt inclusions superimposed on the time evolution of Sr isotopic composition of sea water. **a**, Seawater evolution for the past 2 Gyr (ref. 22). **b**, Detailed Sr evolution for the past 1 Gyr (refs 19, 22). The green diamond indicates seawater composition at 90 Myr (ref. 19) corresponding to the age of lithosphere under the Big Island, Hawaii. The horizontal dashed line represents the pooled average $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 21 analyses of two UDM inclusions. The dotted line and heavy black line indicate the upper and lower limits of range of two standard errors of the average $^{87}\text{Sr}/^{86}\text{Sr}$, respectively. We suggest that the radiogenic Sr in UDMs is derived from ancient sea water in the Mauna Loa source (see the text). The blue field represents the acceptable age range for the recycled seawater component. We reject ages marked by the red field because we have ruled out recent seawater contamination (see the text). We consider the lower limit of the shown range ($^{87}\text{Sr}/^{86}\text{Sr} = 0.7075$, indicated on both panels) to be a minimum estimate for the composition of the seawater component in the Mauna Loa source, because its isotopic composition must have been diluted by less radiogenic components from the original unaltered basalt and source peridotite. The most likely age of recycled seawater Sr is in the range 200–550 Myr, but it may be extended to 650 Myr (ref. 23).

60:40, respectively^{27,28} (Supplementary information). From one-half to two-thirds of this reaction pyroxenite is derived from high-silica melt from recycled crust in the form of eclogite, with the other part being peridotite²⁷. In total this yields 60–70% peridotite component in the source of Mauna Loa magma. However, because ⁸⁷Sr/⁸⁶Sr in the UDMs is nearly at the maximum possible value for sea water¹⁹, this implies that most of the Sr, as well as other incompatible elements, came from the recycled crustal component. This leaves very little room for peridotite-derived incompatible elements, and thus requires that the peridotitic component was severely depleted in these elements, including Sr, so that it did not significantly affect the incompatible element budget of the final melt (see Methods and Supplementary Information for the quantitative model). Alternatively, peridotite could be moderately depleted in incompatible elements, but it would contain radiogenic Sr resulting from seawater alteration, and this would represent part of the recycled oceanic lithosphere.

Our interpretation of the UDM compositions implies an age for the recycled component in the Mauna Loa source of between 200–650 Myr, much younger than previously suggested source differentiation ages of deep mantle plumes^{2–5,7,8}. Such a young age is consistent with a timescale of general mantle circulation with an average rate of about 1–3 cm yr^{−1}, assuming that subducted crust was delivered to the core–mantle boundary at 2,900 km depth, and the Hawaiian plume rises from that depth.

METHODS SUMMARY

The picritic sample K97-15b from Puu Wahi scoria cone (age 910 yr), Mauna Loa volcano, Hawaii, contains euhedral olivine grains (Fo 89–82) with naturally quenched melt inclusions and matrix glass.

Major and some trace elements in melt inclusions, glass and host olivines were determined by electron probe microanalysis on a Jeol JXA 8200 SuperProbe Electron Probe Microanalyser at the Max Planck Institute for Chemistry (MPIC) Mainz, Germany, with a 2σ relative error of 1–2%. The compositions of olivine and contents of Cl, S, Ni and Cr in inclusions were analysed by following a special procedure that allows precision and accuracy of 20–30 μg g^{−1} (2σ error) for Ni, Ca, Mn, Al, Ti, Cr and Co, 0.02 mol% for the forsterite component in olivine²⁸, 4 μg g^{−1} for Cl and 20 μg g^{−1} for S in glass.

LA-ICP-MS was used to determine trace elements in glasses of melt inclusions on an Element-2, Thermo Scientific mass spectrometer with a UP-213 New Wave Research solid-phase laser at MPIC, with a precision and accuracy of about 10–15% (2 relative standard deviations (r.s.d.)). Boron concentrations were measured by secondary-ion mass spectrometry on a Cameca Ims-3f ion probe at MPIC with an accuracy and precision of about 20% (2 r.s.d.).

In situ Sr and Pb isotope analyses of melt inclusions were performed at MPIC with the ICP mass spectrometer Thermo Element-2 and the New Wave UP 193 laser ablation system with a precision and accuracy of about 0.02–0.10% (1 r.s.d.) for Sr (new data and ref. 29) and 0.1% ²⁰⁸Pb/²⁰⁶Pb and ²⁰⁷Pb/²⁰⁶Pb (ref. 30) for one-to-three-spot analyses of melt inclusions.

Quantitative modelling suggests that highly radiogenic Sr and the chemical composition of Mauna Loa UDMs can result from recycling, melting and reacting of depleted oceanic crust altered by sea water and depleted mantle peridotite.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 July 2010; accepted 17 June 2011.

Published online 10 August 2011.

- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- Chase, C. G. Oceanic island Pb: two-stage histories and mantle evolution. *Earth Planet. Sci. Lett.* **52**, 277–284 (1981).
- McKenzie, D. *et al.* Source enrichment processes responsible for isotopic anomalies in oceanic island basalts. *Geochim. Cosmochim. Acta* **68**, 2699–2724 (2004).
- Sun, S. S. & Hanson, G. N. Origin of Ross Island basanitoids and limitations upon the heterogeneity of mantle sources for alkali basalts and nephelinites. *Contrib. Mineral. Petrol.* **52**, 77–106 (1975).
- Tatsumoto, M. Isotopic composition of lead in oceanic basalt and its implication to mantle evolution. *Earth Planet. Sci. Lett.* **38**, 63–87 (1978).
- Farnetani, C. G. & Hofmann, A. W. Dynamics and internal structure of a lower mantle plume conduit. *Earth Planet. Sci. Lett.* **282**, 314–322 (2009).
- Brandon, A. D., Graham, D. W., Weight, T. & Gautason, B. ¹⁸⁶Os and ¹⁸⁷Os enrichments and high-³He/⁴He sources in the Earth's mantle: evidence from Icelandic picrites. *Geochim. Cosmochim. Acta* **71**, 4570–4591 (2007).

- Sobolev, A. V., Hofmann, A. W., Brüggmann, G., Batanova, V. G. & Kuzmin, D. V. A quantitative link between recycling and osmium isotopes. *Science* **321**, 536 (2008).
- Sun, W. D., Bennett, V. C. & Kamenetsky, V. S. The mechanism of Re enrichment in arc magmas: evidence from Lau Basin basaltic glasses and primitive melt inclusions. *Earth Planet. Sci. Lett.* **222**, 101–114 (2004).
- Sobolev, A. V. Melt inclusions in minerals as a source of principal petrological information. *Petrology* **4**, 209–220 (1996).
- Saal, A. E., Hart, S. R., Shimizu, N., Hauri, E. H. & Layne, G. D. Pb isotopic variability in melt inclusions from oceanic island basalts, Polynesia. *Science* **282**, 1481–1484 (1998).
- Jackson, M. G. & Hart, S. R. Strontium isotopes in melt inclusions from Samoan basalts: implications for heterogeneity in the Samoan plume. *Earth Planet. Sci. Lett.* **245**, 260–277 (2006).
- Sobolev, A. V., Hofmann, A. W. & Nikogosian, I. K. Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lavas. *Nature* **404**, 986–990 (2000).
- Sobolev, A. V. & Shimizu, N. Ultra-depleted primary melt included in an olivine from the Mid-Atlantic Ridge. *Nature* **363**, 151–154 (1993).
- Hofmann, A. W. in *Treatise on Geochemistry* Vol. 2 (eds Holland, H. D. & Turekian, K. K.) 61–101 (Elsevier, 2003).
- Coggon, R. M., Teagle, D. A. H., Smith-Duque, C. E., Alt, J. C. & Cooper, M. J. Reconstructing past seawater Mg/Ca and Sr/Ca from mid-ocean ridge flank calcium carbonate veins. *Science* **327**, 1114–1117 (2010).
- Muinos, S. B. *et al.* New constraints on the Pb and Nd isotopic evolution of NE Atlantic water masses. *Geochim. Geophys. Geosyst.* **9**, Q02007 (2008).
- Staudigel, H., Plank, T., White, W. M. & Schmincke, H. U. in *SUBCON: Subduction from Top to Bottom* Vol. 96 (eds Bebout, G. E. & Kirby, S. H.) 19–38 (American Geophysical Union, 1996).
- Veizer, J. *et al.* Sr-87/Sr-86, delta C-13 and delta O-18 evolution of Phanerozoic seawater. *Chem. Geol.* **161**, 59–88 (1999).
- Staudigel, H., Hart, S. R. & Richardson, S. H. Alteration of the oceanic crust: Processes and timing. *Earth Planet. Sci. Lett.* **52**, 311–327 (1981).
- Marshall, H. R., Altherr, R. & Rupke, L. Squeezing out the slab—modelling the release of Li, Be and B during progressive high-pressure metamorphism. *Chem. Geol.* **239**, 323–335 (2007).
- Shields, G. & Veizer, J. Precambrian marine carbonate isotope database: version 1.1. *Geochim. Geophys. Geosyst.* **3**, 1031 (2002).
- Halverson, G. P., Dudas, F. O., Maloof, A. C. & Bowring, S. A. Evolution of the ⁸⁷Sr/⁸⁶Sr composition of Neoproterozoic seawater. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **256**, 103–129 (2007).
- Ross, K. & Elthon, D. Cumulates from strongly depleted mid-ocean-ridge basalt. *Nature* **365**, 826–829 (1993).
- Benoit, M., Ceuleneer, G. & Polvé, M. The remelting of hydrothermally altered peridotite at mid-ocean ridges by intruding mantle diapirs. *Nature* **402**, 514–518 (1999).
- Abouchami, W., Galer, S. J. G. & Hofmann, A. W. High precision lead isotope systematics of lavas from the Hawaiian Scientific Drilling Project. *Chem. Geol.* **169**, 187–209 (2000).
- Sobolev, A. V., Hofmann, A. W., Sobolev, S. V. & Nikogosian, I. K. An olivine-free mantle source of Hawaiian shield basalts. *Nature* **434**, 590–597 (2005).
- Sobolev, A. V. *et al.* The amount of recycled crust in sources of mantle-derived melts. *Science* **316**, 412–417 (2007).
- Jochum, K. P., Stoll, B., Weis, U., Kuzmin, D. V. & Sobolev, A. V. *In situ* Sr isotopic analysis of low Sr silicates using LA-ICP-MS. *J. Anal. At. Spectrom.* **24**, 1237–1243 (2009).
- Jochum, K. P., Stoll, B., Herwig, K. & Willbold, M. Improvement of *in situ* Pb isotope analysis by LA-ICP-MS using a 193 nm Nd:YAG laser. *J. Anal. At. Spectrom.* **21**, 666–675 (2006).
- McDonough, W. F. & Sun, S. S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. T. Anderson for providing the Puu-Wahi sample, N. Groschopf for help in managing the electron probe microanalyser, A. Yasevich and O. Kuzmina for sample preparation, and G. Wörner, N. Arndt and F. Holtz for discussions. This study was funded by an Agence Nationale de la Recherche, France, Chair of Excellence grant (ANR-09-CEXC-003-01) to A.V.S.. Partial support by a Gauss Professorship in Göttingen University, Germany, the Russian Foundation for Basic Research (09-05-01193a), a Russian President grant for leading Russian scientific schools (H111-3919.2010.5), and Earth Sciences Department of Russian Academy grants to A.V.S. are also acknowledged. This is Lamont Doherty Earth Observatory contribution 7479.

Author Contributions A.V.S. designed the project. A.V.S. and A.W.H. conceived the interpretation and the model and wrote the paper. K.P.J. developed the analytical methods for isotope measurements by LA-ICP-MS. D.V.K. processed samples. D.V.K. and B.S. took the measurements. All authors contributed intellectually to the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.V.S. (alexander.sobolev@ujf-grenoble.fr).

METHODS

Samples. The sample K97-15b has been collected by Alfred Anderson at Puu Wahi, from a scoria cone of 910-year-old picritic lava situated at about 3,000 m elevation on the northeast rift zone of Mauna Loa volcano, Hawaii. The melt inclusions hosted by euhedral olivine grains (Fo 89–82) of millimetre size were naturally quenched during eruption and formed fresh glass and small (a few volume per cent) low-density shrinkage bubbles (Supplementary Fig. 1).

Isotope analyses. *In situ* Sr and Pb isotope analyses of melt inclusions were performed at the Max Planck Institute for Chemistry in 2006, with the ICP mass spectrometer Thermo Element-2 and the New Wave UP 193 laser ablation system (wavelength 193 nm, energy density 4 J cm^{-2} , spot sizes $50 \mu\text{m}$ (Sr) and $75\text{--}100 \mu\text{m}$ (Pb), pulse repetition rate 10 Hz)^{29,30}. In addition, Sr was remeasured in October 2010 and February 2011 after improvements of the detection system, to reaffirm and improve the overall reliability of the Sr isotope results.

The precision and accuracy of our LA-ICP-MS method in measuring $^{87}\text{Sr}/^{86}\text{Sr}$ ratios was constrained by analysing well-documented reference glasses (see Supplementary Table 1). The external precision of data obtained in 2006 was reported in ref. 29 as the standard deviation (s.d.) of sample measurements and is a strong function of Sr content of glasses (see Supplementary Fig. 6). From this equation (obtained for the reference glasses) one can predict s.d. = 0.0010 for UDMs containing 65 p.p.m. Sr. For the nine individual UDM measurements in 2006 the corresponding predicted standard error of the mean (s.e.m.) is 0.0003. These predicted values are slightly smaller than the actually observed s.d. of 0.0016 and 0.0005 for UDMs. This result is to be expected because reference glasses are significantly larger in all three dimensions than the measured inclusions and generally yield a somewhat larger number of total counts, yielding slightly higher precision.

To substantiate our results and further improve precision we reanalysed (in October 2010) the Sr isotope ratios of 21 melt inclusions, including both UDMs, using a significantly upgraded LA-ICP-MS system providing a combined effect of an improved stability and higher sensitivity. The improvements include a new multiplier system, a new ion detection unit and front-end computer of the ICP mass spectrometer, and new adjustment of the laser including cleaning of the valves and tubing in the laser ablation system. These technical improvements led to a laser fluence that was higher (about 10 J cm^{-2} in comparison with about 4 J cm^{-2}) and more uniform (decrease to about 2% during 4 h of operation, from about 25% before the improvements), higher counting rates (by about twofold) and lower blanks (for example, for ^{88}Sr about 70 c.p.s. in comparison with 300 c.p.s.²⁹).

The higher precision and accuracy of the new data, for low-Sr samples, is clearly shown by new analyses of the reference glasses (Supplementary Fig. 6 and Supplementary Table 1), performed concurrently with those of the inclusions. Note that all natural reference glasses yield means of measured $^{87}\text{Sr}/^{86}\text{Sr}$ ratios within one standard error of the reference value, which proves that the standard error is an appropriate measure of both precision and accuracy. The predicted maximum precision (and accuracy) of $^{87}\text{Sr}/^{86}\text{Sr}$ ratios for the new measurements of UDM inclusions, estimated from these data, is s.d. = 0.0004 for an individual measurement and s.e.m. = 0.0001 for the mean of 12 measurements. A significant advance in precision of UDMs and reference glasses with similar low Sr contents is evident from Supplementary Tables 1 and 3. All reanalysed melt inclusions match the older data set very well but with better precision (see Supplementary Fig. 7). The new data for 12 individual measurements of the UDM yield $^{87}\text{Sr}/^{86}\text{Sr} = 0.7082 \pm 0.0006$ (two s.e.m.). If we pool the old and the new data, we have 21 measurements yielding a virtually identical mean value of $^{87}\text{Sr}/^{86}\text{Sr} = 0.7081 \pm 0.0006$ (two s.e.m.). Although the precision of the new data are somewhat better than that of the older analyses, we use the full data set of 21 analyses because the two data sets, separated by a time gap of four years, show excellent consistency.

Analysis of major and trace elements. Major and trace elements were determined by electron probe microanalysis on a Jeol JXA 8200 SuperProbe Electron Probe Microanalyser at MPIC, Mainz, Germany. Major-element abundances in glasses were measured at an accelerating voltage of 15 kV and a beam current of 12 nA with a reference sample of natural basaltic glass USNM111240/52 (VG2)³² with a relative error of 1–2%. The compositions of olivines and Cl, S, Ni and Cr in glasses were analysed at an accelerating voltage of 20 kV and a beam current of 300 nA, by following a special protocol²⁸ that allowed 20–30 p.p.m. (2σ error) precision and accuracy for Ni, Ca, Mn, Al, Ti, Cr and Co, and 0.02 mol% for the forsterite component in olivine and 20 p.p.m. for Cl and S in glass.

Trace elements in glasses of melt inclusions were determined by LA-ICP-MS on an Element-2, Thermo Scientific mass spectrometer with a UP-213 New Wave Research solid-phase laser at MPIC, with reference to the KL-2G and NIST 612 standard glasses³³ (see <http://georem.mpch-mainz.gwdg.de>). Ca was used as a reference element. The typical conditions were: laser diameter 60–80 μm , energy density about 7 J cm^{-2} , pulse repetition rate 10 Hz and ablation

time 60–80 s. The element abundances were determined with 2σ errors of no more than 5% and 10% for concentrations above 1 p.p.m. and ~ 0.1 p.p.m., respectively.

Boron content in glasses of melt inclusions was measured by secondary-ion mass spectrometry on a Cameca Ims3F ion microanalyser at MPIC with reference to the KL-2G and NIST 612 standard glasses by following the protocol¹⁰ with a relative error commonly within 10%. The detection limit for B, estimated from ^{11}B intensity on the host olivine, was below 0.02 p.p.m.

Modelling. According to our model, the parental Hawaiian melts are mixtures of melts produced by the melting of olivine-free reaction pyroxenite and peridotite in which the proportion of pyroxenite-derived melt (X_{PX}) is estimated from the olivine composition^{27,28}. The pyroxenitic source is generated by reaction between high-Si, eclogite-derived melt and peridotite, with the proportion of reactive melt (ϕ) similar to the original amount of olivine in peridotite²⁷. The other variables that constrain the composition of the final melt are: degrees of melting of eclogite (F_{E}) and peridotite (F_{PE}), their chemical, isotopic and phase compositions, the degree of melting of pyroxenite (F_{PX}), melting reactions, and partition coefficients between melt and crystals. Mass conservation requires the following relations between the contents of bulk Sr, ^{86}Sr and ^{87}Sr in the final mantle-derived, primary melt (PM), peridotite (PE) and eclogite (E) for a batch melting process:

$$i_{\text{SrPM}} = i_{\text{SrPE}} \times \left[\frac{X_{\text{PX}} \times (1 - \phi)}{F_{\text{PX}} + (1 - F_{\text{PX}}) \times K_{\text{PX}}^{\text{Sr}}} + \frac{(1 - X_{\text{PX}})}{F_{\text{PE}} + (1 - F_{\text{PE}}) \times K_{\text{PE}}^{\text{Sr}}} \right] + i_{\text{SrE}} \times \frac{\phi \times X_{\text{PX}}}{(F_{\text{PX}} + (1 - F_{\text{PX}}) \times K_{\text{PX}}^{\text{Sr}}) \times (F_{\text{E}} + (1 - F_{\text{E}}) \times K_{\text{E}}^{\text{Sr}})} \quad (1)$$

where $i = 86$ or 87 , nd (bulk Sr) $K_{\text{PX}}^{\text{Sr}}$, $K_{\text{PE}}^{\text{Sr}}$ and K_{E}^{Sr} are bulk distribution coefficients for Sr or its isotopes between crystal phases and melt for pyroxenite, peridotite and eclogite, respectively.

$$^{86}\text{Sr}_j = \frac{\text{Sr}_j}{^{87}\text{Sr}_j/^{86}\text{Sr}_j + (^{84}\text{Sr}_j + ^{86}\text{Sr}_j + ^{88}\text{Sr}_j)/^{86}\text{Sr}_j} \quad (2)$$

where $j = \text{PM, PE or E}$, and the ratio of unradiogenic stable isotopes ($^{84}\text{Sr}_j + ^{86}\text{Sr}_j + ^{88}\text{Sr}_j$)/ $^{86}\text{Sr}_j = 9.43205$.

Equation (1) for Sr, ^{87}Sr and ^{86}Sr and equation (2) for primary melt, eclogite and peridotite give six equations with nine unknowns (Sr_{PE} , $^{87}\text{Sr}_{\text{PE}}$, $^{86}\text{Sr}_{\text{PE}}$, Sr_{PM} , $^{87}\text{Sr}_{\text{PM}}$, $^{86}\text{Sr}_{\text{PM}}$, Sr_{E} , $^{87}\text{Sr}_{\text{E}}$ and $^{86}\text{Sr}_{\text{E}}$). To resolve these equations one needs to constrain three unknowns. The seawater model presented in this paper allows a maximum $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of recycled eclogite of 0.7090 (first constraint), while the measured composition of the UDMs requires the minimum $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of the final melt to be 0.7075 (second constraint; see Fig. 4). The composition of the UDMs also allows the composition of primary UDM to be estimated by reversing olivine fractionation up to equilibrium with the most magnesian olivine for Mauna Loa (Fo 90.7; ref. 27). This yields $\text{Sr}_{\text{PM}} = 51$ p.p.m. Sr in the primary UDM melt (third constraint).

Solving mass balance equations (1) and (2) for the above constraints and using reasonable values of ϕ (0.60–0.65), F_{E} (0.5), F_{PE} (0.10–0.15), F_{PX} (0.45), X_{PX} (0.60), $K_{\text{PX}}^{\text{Sr}} = 0.05$, $K_{\text{PE}}^{\text{Sr}} = 0.01$ and $K_{\text{E}}^{\text{Sr}} = 0.15$ (see the discussion on these parameters in ref. 27) gives a relation between contents and isotope composition of Sr in the peridotite component (Supplementary Fig. 9). Assuming a lower limit for the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of peridotite of 0.7021 (extreme depletion) yields a maximum Sr content of peridotite of about 3 p.p.m., which fulfils the conditions stated above. Thus, as expected, the peridotite source should be more depleted in Sr than average depleted MORB mantle with a Sr content of 7.66 p.p.m.³⁴. For a lower content or more radiogenic Sr in peridotite, the $^{87}\text{Sr}/^{86}\text{Sr}$ in the recycled crust would be lower than 0.7090. It approaches 0.7075 (minimum allowed by the observed UDM composition) when there is no Sr in peridotite or the ratio $^{87}\text{Sr}/^{86}\text{Sr}$ in peridotite is 0.7075.

The above model is somewhat oversimplified because it uses a batch melting process and considers fixed melting phase proportions in the form of constant bulk distribution coefficients. To take into account a more realistic, aggregated critical melting process and include changing phase proportions, we further use the conceptually similar but more advanced model with melting reactions and distribution coefficients explained in ref. 27. Using the above estimates for maximum Sr content in peridotite we chose for our quantitative modelling the composition of depleted abyssal harzburgite from the database³⁵ with a Sr content of 2.73 p.p.m. (model 1), and a restite after 2% of near fractional melting of depleted MORB mantle³⁴ with a Sr content of 1.96 p.p.m. (model 2). The results of modelling are shown in Supplementary Fig. 10 and Supplementary Table 4.

The composition of recycled crust was calculated to match the trace-element composition and minimum $^{87}\text{Sr}/^{86}\text{Sr} = 0.7075$ ratio of UDM primary melt. For $^{87}\text{Sr}/^{86}\text{Sr} = 0.7021$ of peridotite, the results for recycled crust are $^{87}\text{Sr}/^{86}\text{Sr} = 0.7086$

(model 1) and 0.7083 (model 2). The estimated trace-element compositions of recycled crust correspond to highly depleted oceanic crust and are reasonably close to reported compositions of depleted MORBs^{36,37} (Supplementary Fig. 10a). The compositions of peridotites correspond to depleted abyssal harzburgite or depleted lherzolite (Supplementary Fig. 10a). We conclude that the highly radiogenic Sr and the chemical composition of Mauna Loa ultra-depleted melts can be produced by recycling, melting and reacting of depleted oceanic crust altered by sea water and depleted mantle peridotite. This peridotite may be also a part of the recycled oceanic lithosphere.

We further investigate an alternative possibility for the production of highly radiogenic Sr in ultra-depleted melt by the involvement of small amounts of continental sediments with an extremely high $^{87}\text{Sr}/^{86}\text{Sr}$ ratio. We use the same model as model 1 but for $^{87}\text{Sr}/^{86}\text{Sr} = 0.7025$ of recycled oceanic crust and variable amounts of sediment with the most extreme $^{87}\text{Sr}/^{86}\text{Sr} = 0.7349$ (from Sumatra³⁸). This yields an amount of 1.6% of such sediment in the recycled component to match $^{87}\text{Sr}/^{86}\text{Sr} = 0.7075$ of the final melt. However, this substantially increases the concentration of highly incompatible elements such as Th and La of the final melt and significantly lowers its Nb/Th ratio (Supplementary Fig. 10d and Supplementary Table 4). We conclude that recycled continental material, even

with extremely radiogenic Sr, cannot produce the highly incompatible trace-element pattern of Mauna Loa UDMs.

32. Jarosevich, E. J., Nelen, J. A. & Norberg, J. A. Reference sample for electron microprobe analysis. *Geostand. Newsl.* **4**, 43–47 (1980).
33. Jochum, K. P. *et al.* The preparation and preliminary characterization of eight geological MPI-DING reference glasses for in-situ microanalysis. *Geostand. Newsl.* **24**, 87–133 (2000).
34. Workman, R. K. & Hart, S. R. Major and trace element composition of the depleted MORB mantle (DMM). *Earth Planet. Sci. Lett.* **231**, 53–72 (2005).
35. Niu, Y. L. Bulk-rock major and trace element compositions of abyssal peridotites: implications for mantle melting, melt extraction and post-melting processes beneath mid-ocean ridges. *J. Petrol.* **45**, 2423–2458 (2004).
36. Saal, A. E., Hauri, E. H., Langmuir, C. H. & Perfit, M. R. Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* **419**, 451–455 (2002).
37. Bach, W., Peucker-Ehrenbrink, B., Hart, S. R. & Blusztajn, J. S. Geochemistry of hydrothermally altered oceanic crust: DSDP/ODP Hole 504B—implications for seawater–crust exchange budgets and Sr- and Pb-isotopic evolution of the mantle. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2002GC000419 (2003).
38. Plank, T. & Langmuir, C. H. The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chem. Geol.* **145**, 325–394 (1998).

Structure of the membrane domain of respiratory complex I

Rouslan G. Efremov^{1†} & Leonid A. Sazanov¹

Complex I is the first and largest enzyme of the respiratory chain, coupling electron transfer between NADH and ubiquinone to the translocation of four protons across the membrane. It has a central role in cellular energy production and has been implicated in many human neurodegenerative diseases. The L-shaped enzyme consists of hydrophilic and membrane domains. Previously, we determined the structure of the hydrophilic domain. Here we report the crystal structure of the *Escherichia coli* complex I membrane domain at 3.0 Å resolution. It includes six subunits, NuoL, NuoM, NuoN, NuoA, NuoJ and NuoK, with 55 transmembrane helices. The fold of the homologous antiporter-like subunits L, M and N is novel, with two inverted structural repeats of five transmembrane helices arranged, unusually, face-to-back. Each repeat includes a discontinuous transmembrane helix and forms half of a channel across the membrane. A network of conserved polar residues connects the two half-channels, completing the proton translocation pathway. Unexpectedly, lysines rather than carboxylate residues act as the main elements of the proton pump in these subunits. The fourth probable proton-translocation channel is at the interface of subunits N, K, J and A. The structure indicates that proton translocation in complex I, uniquely, involves coordinated conformational changes in six symmetrical structural elements.

Complex I (NADH:ubiquinone oxidoreductase) is one of the largest known membrane proteins. It provides about 40% of the proton flux during proton-motive force generation for the synthesis of ATP by the mitochondrial respiratory chain^{1–5}. Many mutations in subunits of complex I have been associated with human neurodegenerative diseases^{5,6}. Complex I is a major source of reactive oxygen species in mitochondria: these can damage mitochondrial DNA and are considered to be causative agents in Parkinson's disease⁷ and ageing⁸. Mitochondrial complex I consists of 45 subunits (980 kDa in total)⁹. The simpler prokaryotic enzyme usually consists of 14 'core' subunits (about 550 kDa in total) that are conserved from bacteria to humans^{1,2,5,10}. It contains equivalent redox components and has a similar L-shaped structure to that of mitochondrial complex I (refs 2, 5, 11). The high degree of sequence conservation of core subunits indicates that the mechanism of complex I is probably the same throughout all species.

The complete atomic structure of this molecular machine is currently unknown. Structures of the *Thermus thermophilus* complex I hydrophilic domain^{12,13} have established the electron-transfer pathway from NADH to the primary electron acceptor flavin mononucleotide, and then through seven conserved iron-sulphur clusters to the putative quinone binding site at the interface with the membrane domain (which lacks covalently bound prosthetic groups). The three largest hydrophobic subunits of complex I, NuoL, M and N (*E. coli* nomenclature), are homologous to each other and to Na⁺/H⁺ antiporter complex (Mrp) subunits^{14,15}. They are likely to participate in proton translocation. We previously determined the arrangement of α -helices in the membrane domain of *E. coli* complex I, but the limited resolution (3.9 Å) prevented sequence assignment¹¹.

Understanding how the transfer of two electrons from NADH to ubiquinone is coupled to the translocation of four protons (the current consensus value) across the membrane remains a major question in complex I research^{1–3,5}. 'Direct' (redox-driven) and 'indirect'

(conformation-driven) mechanisms have been proposed^{2,3,5,16}. Recently published structures revealed an unexpected structural element, helix HL, extending along nearly the entire length of the membrane domain^{11,17}: this may act as a connecting element, coordinating conformational changes¹¹. A greater understanding requires a knowledge of the atomic structure of the membrane domain.

Structure determination and overall architecture

A new P1 crystal form was obtained with protein purified in the detergent cymal-7. After extensive optimization, the diffraction extended to 2.7 Å in the best direction. The crystallized fragment contains six subunits, NuoL, M, N, A, J and K (222 kDa in total), but lacks NuoH, which dissociates readily from the complex^{11,18}. Multiwavelength anomalous diffraction (MAD) data sets were collected from SeMet-derivatized protein (Supplementary Table 1). Multi-crystal density averaging and modification led to experimental electron density maps of excellent quality (Supplementary Fig. 1). The model was refined to 3.0 Å resolution with $R = 23.2\%$ and $R_{\text{free}} = 28.3\%$ (Supplementary Table 1), and includes 1,952 residues (out of 2,038, Supplementary Table 2). A cymal-7 molecule, fragments of lipid aliphatic chains and several well-ordered internal water molecules are also included.

The model contains 55 transmembrane helices (TMs) (Fig. 1 and Supplementary Movie 1). Subunits NuoA, J and K form an 11-helix bundle near the interface with the hydrophilic domain, and the antiporter-like subunits NuoN, M, and L follow, as proposed earlier¹¹. The distal subunit NuoL contains a carboxy-terminal extension that starts with TM15, followed by helix HL and ending with TM16, harboured at the interface with subunits NuoJ, K and N. Sequence conservation is high for the membrane core of all subunits (Supplementary Fig. 3), underlining its functional significance. Helix HL is poorly conserved, as might be expected for a mostly mechanical element, with the exception of residues contacting other subunits

¹Medical Research Council Mitochondrial Biology Unit, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK. [†]Present address: Max-Planck Institute for Molecular Physiology, Otto-Hahn Strasse 11, Dortmund 44227, Germany.

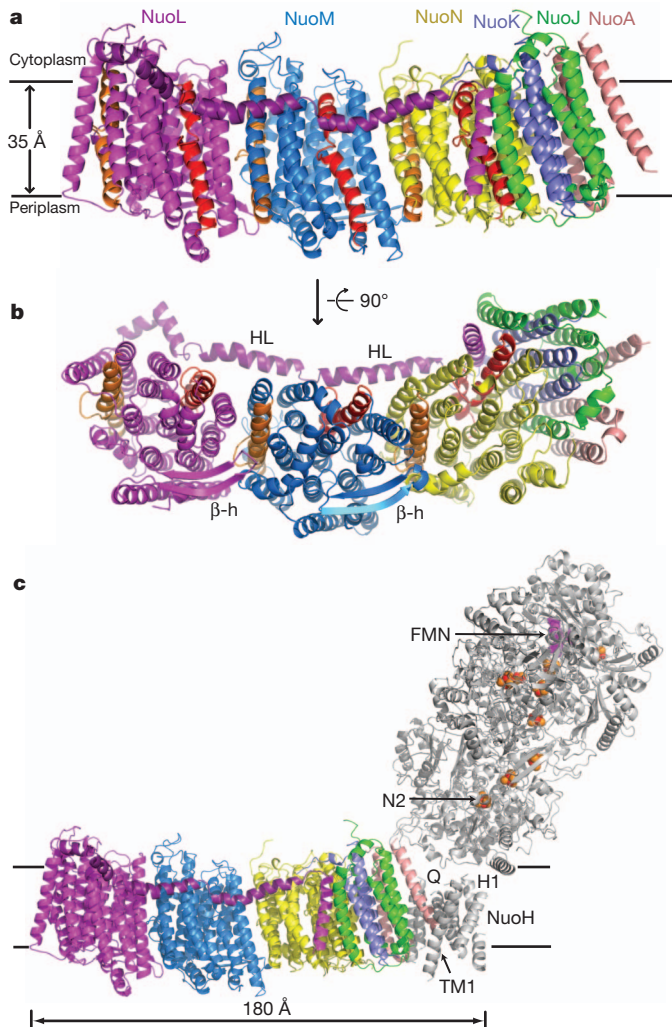


Figure 1 | Architecture of the membrane domain of *E. coli* complex I. Cartoon representation of the atomic model. **a**, Side view in the membrane plane. **b**, View from the periplasm into the membrane. Subunits are coloured as indicated by the labels. Discontinuous helices are shown in red (TM7) and orange (TM12). Connecting elements, helix HL and β -hairpins (β -h), are labelled. The position of the lipid bilayer is estimated as in Supplementary Fig. 4. **c**, Side view of the structure, aligned with the membrane domain of the entire complex I from *T. thermophilus* (PDB code 3M9S)¹¹. Subunits that are not present in the *E. coli* structure are shown in grey. Flavin mononucleotide (FMN) is shown as magenta spheres; iron-sulphur clusters are shown as red-orange spheres, with cluster N2 labelled. Helix TM1 from subunit Nqo8/NuoH, helix H1 from Nqo6/NuoB and the probable quinone-binding site (Q) are indicated.

(Supplementary Fig. 3b), indicating that this helix is similarly arranged across species. The polar residues of amphiphatic helix HI face the cytoplasm and apolar residues point to the membrane (Supplementary Fig. 4), indicating that helix HL resides at the surface of the lipid bilayer. The periplasmic surface of the membrane domain is mostly negatively charged and the cytoplasmic surface is positively charged, in agreement with the positive-inside rule¹⁹. Lipid bilayer boundaries are clearly visualized by the belts of solvent-exposed Tyr and Trp residues²⁰ and by the hydrophobicity of surface-exposed residues (Supplementary Fig. 4).

Fold of antiporter-like subunits

Subunits NuoL, M and N contain 14 structurally conserved helices (Fig. 2a–c, Supplementary Fig. 5 and Supplementary Table 3). The fold differs from the bioinformatic prediction²¹. Sequential transmembrane helices contact each other, except for TM8 and TM9,

and TM13 and TM14. The two discontinuous helices observed previously¹¹ are TM7 and TM12. They are interrupted in the middle of the bilayer by an extended loop of 5–7 residues. The tips of these loops contain a proline that is conserved between all three antiporter-like subunits: M(Pro 239)* in TM7 and M(Pro 399)* in TM12 (the prefix indicates the subunit name; asterisks indicate residues conserved between all three antiporter-like subunits). As in other transporters^{22,23}, such helices are likely to participate in proton or ion transport by introducing some flexibility and charge to the middle of the membrane.

TMs 1, 2 and 14 form a lipid-facing layer at the opposite side of the domain from helix HL and are the least conserved in sequence and in structure (Supplementary Figs 3 and 5). TMs 2 and 3 are connected at the periplasmic surface by an extended β -hairpin. The core of the subunits is formed by the well-conserved helices 4–13. Their arrangement has internal symmetry: TMs 4–8 can be superimposed on TMs 9–13 by a rotation of about 180° along the axis lying in the membrane plane and directed along the long axis of the domain, followed by a shift along this axis (Fig. 2c–f and Supplementary Movie 2). Thus, the two halves of the core, each containing five TMs, are related, uniquely, by symmetry along a pseudo-two-fold screw axis. Symmetry-related sets of helices in transporters may span the membrane in a parallel or antiparallel fashion, but have always been found in face-to-face arrangements, with substrate binding sites located near the symmetry axis at the domain interface²². However, in case of complex I antiporter-like subunits, the interface is formed by face-to-back contact, presenting a novel arrangement of symmetry-related helices. This places discontinuous helices 7 and 12 far apart, so that they cannot interact directly, in contrast to the situation in other transporters. It is likely that antiporter-like subunits have evolved by gene duplication²², even though the sequences of the two inverted domains (Fig. 2c–e) have diverged (Supplementary Table 3).

At the centre of each antiporter-like subunit, TM8 is partly unwound in the middle and contains a kink with no proline residues, similar to the π -bulge in bacteriorhodopsin²⁴. The secondary structure is markedly disrupted near the kink (residues L(253–258), M(259–264) and N(242–247)), indicating flexibility. Such π -bulges (or π -helices²⁵) are normally found at protein functional sites²⁵, indicating a mechanistic importance for the kink in TM8.

Two well-studied, essential charged residues that are conserved from Mrp antiporters to complex I (refs 15, 26) are found in the middle of TM5 (L(Glu 144), M(Glu 144) and N(Glu 133); termed Glu TM5 here), and in the middle of TM7 (L(Lys 229), M(Lys 234) and N(Lys 217); Lys TM7). Lys TM7 is the last residue of the periplasmic half of discontinuous TM7 (TM7a), and its ϵ -amino group points inside the putative proton channel (termed here the first channel), formed by TMs 5–8. Notably, Glu TM5 is wedged at the interface of TMs 5 and 6, and is exposed both to this channel and to the cavity at the interface with the adjacent subunit. The Glu TM5 carboxyl group can interact electrostatically with the Lys TM7 amino group, which is about 5–6 Å away. An invariant Trp from TM7b (M(Trp 243)*) stabilizes the link between the two halves of TM7. The negative helix-dipole charge here may be partly compensated by N(Trp 247), M(His 248) and L(His 254).

The area near the second discontinuous helix, TM12, harbours many conserved polar residues in the middle of the membrane, so is also of clear functional importance. In NuoL and NuoN, an invariant, essential^{127,28} lysine (L(Lys 399) and N(Lys 395)) lies at the beginning of the periplasmic half of TM12 (TM12b) in a position approximately symmetrical to Lys TM7 (Fig. 2d, e). It may interact with, and stabilize, the negatively charged C terminus of TM12a. In NuoM, a similar position is occupied by an invariant Glu 407. Here, the negative dipole moment may be stabilized by the invariant His 322. Notably, Glu 407 is replaced by an invariant lysine in the homologous MrpD subunit of antiporters. The side chain of this Lys or Glu (termed Lys/Glu TM12) is exposed in a large cavity between TMs 10–13. Thus, unexpectedly, a second putative proton channel is present here, related by inverted

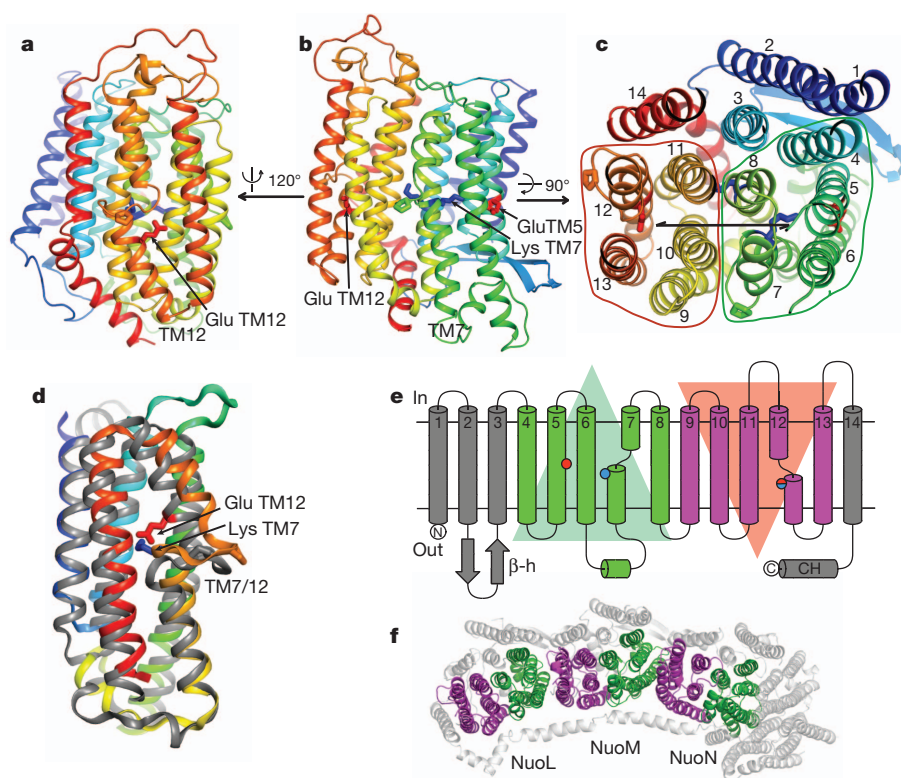


Figure 2 | Fold of antiporter-like subunits. The fold is illustrated using subunit NuoM. **a**, Side view, cytoplasmic side up, centred on TM12. **b**, Side view centred on TM7, coloured blue to red from N to C terminus. Essential charged residues are shown as sticks (Glu TM5 and Lys TM7 from channel 1, Glu TM12 from channel 2 and the connecting M(Lys 265) from TM8). Conserved prolines from intra-helical loops are also shown. **c**, View from the cytoplasm into the membrane. TM helices are numbered. Two inverted repeats are circled and the

pseudo-two-fold symmetry screw axis that relates them is shown. **d**, Overlay of the two symmetry-related domains (TMs 9–13, in grey, over TMs 4–8).

e, Topology diagram of antiporter-like subunits. Two inverted repeats of the conserved core are shown in green and magenta. Crucial charged residues (Glu TM5, Lys TM7 and Lys/Glu TM12) are indicated. **f**, Structural repeats in the membrane domain, seen from the cytoplasm. TMs 4–8 are highlighted in green and TMs 9–13, in magenta.

symmetry to the first channel. Additionally, N(Lys 395) is exposed to the cavity between NuoN and NuoM, and M(Glu 407) is partially exposed to the cavity between NuoM and NuoL, similarly to Glu TM5.

Searches using either complete subunits or the repeating domain revealed no structural analogues in the protein data bank (PDB), confirming that the fold is novel. Homologous subunits in Mrp antiporters, chloroplast Ndh complexes and membrane-bound hydrogenases, not yet structurally characterized, are likely to share the same core fold.

Fold of subunits NuoK, NuoJ and NuoA

In contrast to NuoL, M and N, the smaller subunits interact extensively, with large surface areas buried at interfaces. They also form more hydrogen bonds and salt bridges to their neighbours (Supplementary Table 4), ‘gluing’ together this end of the membrane domain.

NuoK spans the membrane with three linearly arranged α -helices, connected by short loops (Supplementary Fig. 6a). NuoK interacts extensively with NuoN (Fig. 1b), consistent with the presence of the analogous MrpC subunit in antiporters²⁹. The C terminus of NuoK extends between helix HL and NuoN, forming many inter-subunit links. NuoK can be aligned with TMs 4–6 of antiporter-like subunits (and also with TMs 9–11, but with low sequence identities) (Supplementary Table 3). Notably, conserved K(Glu 36) aligns to Glu TM5, indicating that NuoK and the antiporter-like subunits might share an ancestor.

Subunit NuoJ has an unusual non-globular fold: three linearly arranged amino-terminal TM helices border NuoK, and TMs 4 and 5 are separated at the opposite sides of the domain (Supplementary Fig. 6b). Thus, NuoJ interweaves between NuoK, A and N, stabilizing

the complex. TM3 of NuoJ contains in its middle a π -bulge/kink (residues 59–62), similar to TM8 in NuoL, M and N, so this helix is probably flexible and mechanistically important.

Subunit NuoA consists of three linearly arranged TM helices at the edge of the domain, interacting with NuoJ, N and K (Supplementary Fig. 6c). Its outer surface, including the conserved A(Asp 79) in the middle of the membrane, probably interacts with NuoH (Fig. 1c). A conserved loop between TM1 and TM2 is not resolved in the structure and probably interacts with the hydrophilic subunit NuoD¹¹. The probable quinone-binding site is formed at the interface of subunits NuoH, J and A with the hydrophilic domain¹¹ (Fig. 1c).

Interactions between subunits

About half of the total subunit surface area is buried away from the solvent (Supplementary Table 4), consistent with the relative stability of the domain. However, interactions between antiporter-like subunits are not extensive. They mainly involve two helices from each subunit (TM5 and 6 and TM12 and 13), and are mostly hydrophobic, with very few hydrogen bonds (Supplementary Table 4). Crevices between subunits are likely to be filled with lipids, some of which are ordered and modelled here.

Helix HL stabilizes the complex. It is tightly anchored to NuoN, J and K by L(TM16), which also contacts N(TM7). HL contacts NuoM near M(TM7), where HL is unwound, allowing for strong interactions via the backbone and side chains (Fig. 3 and Supplementary Table 4). In a similar fashion, helix HL forms hydrogen bonds to the main body of the domain near TM7 of NuoL and N. These contacts involve conserved M(Lys 173)*, M(Asp 246)* and M(Tyr 317)*, and a semi-conserved residue, M(His 241).

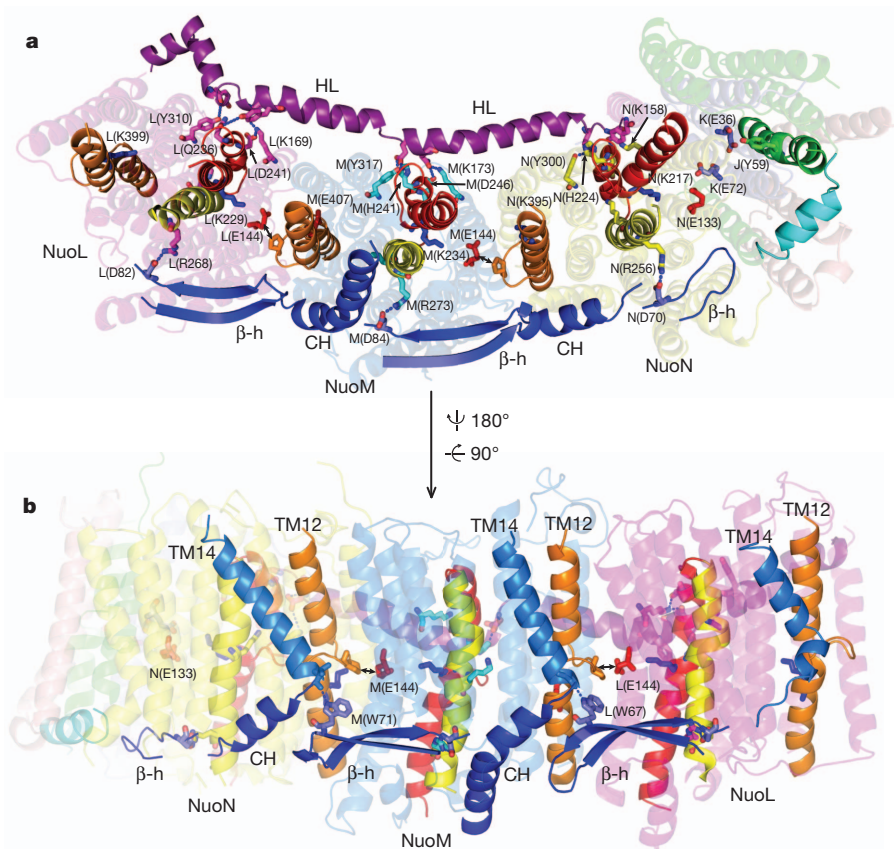


Figure 3 | Connecting elements and interactions between subunits. **a**, View from the periplasm. Helix HL is in purple and the β -hairpin-helix element β H (consisting of hairpins β -h and helices CH) is in blue. The contacting helix from NuoJ (preceding TM5 of NuoJ) is in cyan. Helices that are probably involved in conformational changes are shown in red (TM7 in NuoL, M and N), orange (TM12), yellow (TM8) and green (TM3 in NuoJ). Essential charged residues are shown as sticks and labelled (prefix indicates subunit). Also shown are

conserved residues involved in contacts between antiporter-like subunits and helix HL, as well as between the β -hairpins and TM8. Putative hydrogen bonds are indicated. Potential interactions between Glu TM5 and a conserved proline from the TM12 loop are indicated by arrows. **b**, View in the membrane, with TM14 also shown (light blue). Tryptophan residues forming hydrogen bonds to the backbone at the C terminus of TM14 are indicated.

Notably, an additional element connecting antiporter-like subunits is present on the opposite side of the domain from helix HL. The extended and well-ordered (thus probably rigid) β -hairpins from neighbouring subunits contact each other via a stretch of C-terminal amphipathic helices termed CH (Fig. 3). This notable β -hairpin-helix element (termed the β H element) extends over the entire length of the antiporter-like subunits and contributes to the stability of the complex, because the only polar interactions between NuoM and NuoN are via the tip of the NuoM β -hairpin (there are also similar interactions between NuoM and NuoL; Supplementary Table 4). Predictions of secondary structure indicate that there are β -strands of similar length between TMs 2 and 3 in most other species, so the β -hairpin and second connecting element are likely to be common features.

Proton translocation pathways in subunits NuoL, M and N

The central positions of Lys TM7 and Lys TM12 in both channels, close to the breaks in TM7 and TM12, and the severe effects of their mutations (Supplementary Table 6), indicate that in complex I, unusually, lysine rather than carboxylate residues have a central role in the proton pump (one exception is M(Glu 407)). Examination of potential pathways (Supplementary Discussion, section 1) shows that none of the channels is solvent-accessible on both sides of the membrane.

In the first channel, the cavity surrounding Lys TM7 is about 70 \AA^3 and is lined by 2–5 polar residues (Fig. 4a–c and Supplementary Table 5). It is closed from the periplasm by large hydrophobic residues but is open to the cytoplasm via a network of polar residues. Cavities between

subunits L and M, and M and N, are blocked from the periplasm but may be accessible from the cytoplasm, possibly allowing an additional (but less likely; Supplementary Discussion, section 1) protonation route for Lys TM7, by ‘side-entry’ near Glu TM5.

The cavity in the second channel, near Lys/Glu TM12, is markedly larger and more hydrophilic than that in the first one (up to 200 \AA^3 , ~ 10 polar residues; Fig. 4a–c and Supplementary Table 5). Access to the cytoplasm is closed but the periplasm is accessible via short polar networks.

Many conserved charged and polar residues in the middle of the membrane connect the two channels in each subunit (Fig. 4, Supplementary Fig. 8 and Supplementary Table 5). Notably, in NuoM and NuoN, central positions in this connection are occupied by invariant essential residues^{28,30,31} N(Lys 247) and M(Lys 265), found near the TM8 π -bulge. NuoL also contains an invariant residue, His 254, near this position.

Thus, in all antiporter-like subunits, unusually, two half-closed symmetry-related channels are linked, forming a single continuous proton translocation pathway through each subunit. Consistently, all the observed water molecules are either in the channels or in the connection (Fig. 4). Other pathways seem to be unlikely (Supplementary Discussion, section 2).

The fourth proton translocation pathway

An additional channel is apparent at the interface of subunits NuoN, K, J and A. A cavity of $\sim 130 \text{ \AA}^3$ between NuoN and NuoK can accommodate water molecules coordinated by K(Glu 72), N(Thr 160) and N(Ser 156)

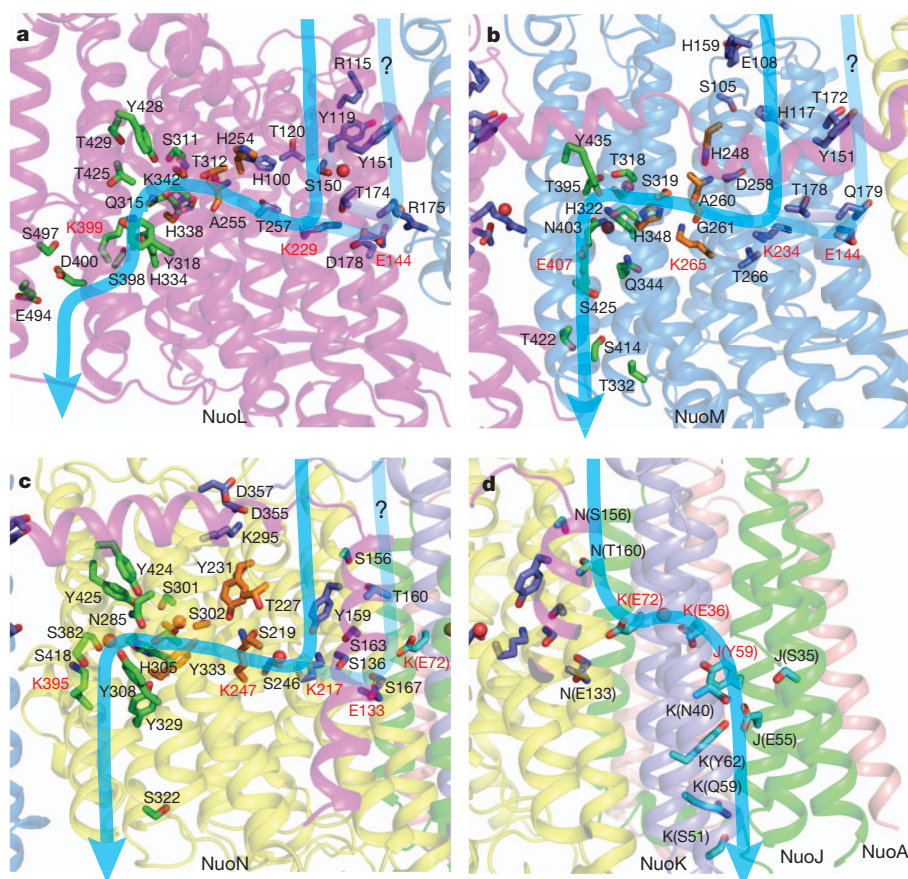


Figure 4 | Proton translocation channels. **a–c**, Putative channels in the antiporter-like subunits. Polar residues lining the channels are shown as sticks, with carbon in dark blue for the first channel, in green for the second channel and in orange for connecting residues. **d**, Channel at the interface of subunits NuonN, K and J, with polar residues in cyan. All views are in a similar orientation to Fig. 1a. Approximate proton translocation paths are indicated by blue

arrows. Additional, less likely, input channels at the interfaces of subunits are paler in colour and labelled '?'. Essential charged residues are labelled in red. Water molecules resolved in the structure are shown as red spheres. Alanine and glycine from the π -bulge in TM8 may coordinate water molecules via exposed backbone carbonyls.

(Fig. 4d). Space for the cavity is provided by an invariant residue, K(Gly76). Access to the periplasm is blocked by large hydrophobic residues but the cytoplasm is accessible with only a short constriction near K(Leu83). Indeed, the structure shows a water molecule in the middle of the bilayer, between the conserved, essential residues^{32,33} K(Glu36) and K(Glu72). From K(Glu36), the continuation of the pathway is apparent at the interface of TMs 2 and 3 from NuokK and NuojJ, reaching the periplasm. Several small cavities here may contain water molecules coordinated by conserved polar residues (Fig. 4d). An alternative 'sideways' pathway from K(Glu36) to the periplasm through invariant A(Glu81) to A(Glu102) can be considered, but is less likely, requiring a more extended network of water molecules.

Mechanism of proton translocation

There is no evidence in the structure for any potential electron carriers in the membrane domain, for the existence of quinone-binding sites in antiporter-like subunits, or for a different role of NuonN (Supplementary Discussion, section 4), as has been discussed in the literature^{34–36}. Structural features combine to indicate strongly that complex I operates purely by a conformation-driven mechanism.

Conformational changes in the hydrophilic domain, which occur upon reduction¹³, can be transmitted to the membrane-domain subunits NuohH, A, J and K (Supplementary Discussion, section 3). Helix HL can relay these changes to TM7 in NuolL, M and N, leading to protonation of Lys TM7 upon closer approach of Glu TM5. Recent studies have confirmed the essential coupling role of this helix^{37,38} (Supplementary Discussion, section 4).

The structure indicates that conformational changes in both half-channels of the antiporter-like subunits are coordinated not only by helix HL, but by several connecting elements, acting in concert to achieve tight coupling of proton translocation events. HL can drive TM7 in the first channel, and, on the opposite side of the domain, the β H element can drive TM12 in the second channel. The two channels can interact both within subunits (via TM8, which contains an invariant connecting Lys or His near its flexible kink) and also between subunits, because the intra-helical loop of TM12 is in close contact with Glu TM5 on the neighbouring subunit. Furthermore, Glu TM5 can also interact electrostatically with Lys/Glu TM12 from a neighbouring subunit, because both residues are exposed to the inter-subunit cavities.

We suggest the following translocation cycle via antiporter-like subunits (Fig. 5), assuming that Lys TM7 is protonated and that Lys/Glu TM12 is deprotonated in the oxidized state (meaning the state of iron-sulphur cluster N2 and a nearby cluster¹³, because most clusters are reduced during turnover³⁹). Conformational changes upon reduction move Glu TM5 away from Lys TM7, forcing lysine to donate its proton into the link between the two half-channels and eventually to Lys/Glu TM12, which becomes protonated owing to local conformational changes, either with or without help from Glu TM5 from the neighbouring subunit. Upon return to the oxidised-state conformation, Glu TM5 moves back, Lys TM7 is protonated from the cytoplasm, reloading the pump, and Lys/Glu TM12 ejects its proton into the periplasm. Overall, three subunits translocate three protons.

In the fourth putative channel, a notable feature indicating conformational coupling is the π -bulge on TM3 of NuojJ, which contains a

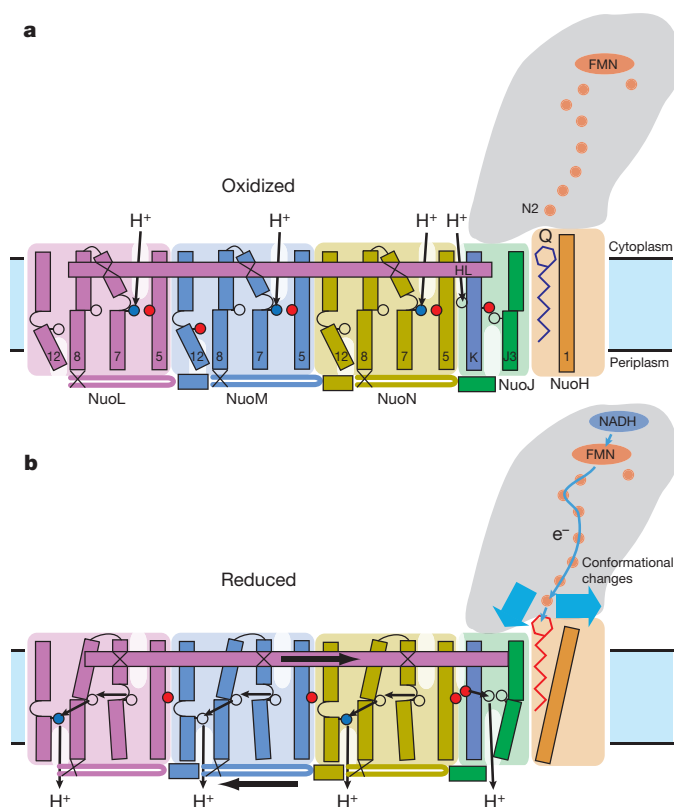


Figure 5 | Proposed mechanism of complex I. **a**, Oxidized state. **b**, Reduced state. Conformational coupling between electron transfer and proton translocation is mediated by helix HL (cytoplasmic side) and the β H element (periplasmic side). TM helices are numbered. Crucial charged residues (Glu TM5, Lys TM7, Lys/Glu TM12 and Lys/His TM8 from Nuol, M and N, as well as K(Glu 72) and K(Glu 36), interacting with J(Tyr 59)) are indicated by red or empty circles (Glu) and empty or blue circles (Lys) for unprotonated or protonated residues, respectively. In Nuol, M and N, Lys TM7 from the first half-channel is assumed to be protonated in the oxidized state. Upon reduction, it donates its proton to the connecting Lys/His TM8 and then on to Lys/Glu TM12 from the second half-channel. Lys/Glu TM12 ejects its proton into periplasm upon return from the reduced to the oxidized state. A fourth proton per cycle is translocated at the interface of Nuon, K and J.

residue that is invariant and important for activity^{40,41}, J(Tyr 59), forming a hydrogen bond to K(Glu 36), a central residue of this channel. TM3 is the most conserved helix in Nuoj (Supplementary Figs 3a and 8) and is a hotspot for human mitochondrial disease mutations (Supplementary Table 7). Conformational changes, propagating from the nearby hydrophilic domain, can move the flexible TM3 of Nuoj and J(Tyr 59), resulting in the protonation/deprotonation of K(Glu 36) and the translocation of the fourth proton (Fig. 5). Because the four channels that are present in the structure account for the measured H^+/e^- stoichiometry⁴², any additional direct coupling, as proposed recently^{43,44}, is unlikely.

Our knowledge of the structure of the core subunits of complex I is now nearly complete (except for subunit Nuoh). The effects of many mutations, including those causing human diseases, can now be understood (Supplementary Tables 6–8). This molecular machine has one of the most intricate architectures among known protein structures. The redox energy of NADH, binding at the tip of the hydrophilic domain, is eventually used in the membrane domain, at distances of up to ~ 300 Å away. Our previously published analogy of the ‘steam engine of the cell’ is emphasized by the presence of several connecting elements, including helix HL, which are powered by the redox energy of electron transfer and which drive six symmetrical structural domains, plus one asymmetrical domain, resulting in a very effective proton pump.

METHODS SUMMARY

The membrane domain of *E. coli* complex I was purified as previously described¹¹, except that the detergent cymal-7 was used. Protein was crystallized in the presence of lipids using polyethylene glycol (PEG) 4000 and sodium tartrate as the precipitant at pH 4.8, with the additives 0.15% sodium cholate and 0.1 M NDSB-256 (non detergent sulfobetaine-256). Diffraction was improved by a slow increase of PEG 4000 concentration to 30%, by dialysis at 4 °C.

Diffraction data were collected from cryo-cooled crystals (100 K) on beamline ID29 at the European Synchrotron Radiation Facility (ESRF) and X06SA at the Swiss Light Source (SLS). Crystallographic phases were obtained using selenomethionine-labelled protein. Electron density was improved by multi-crystal averaging and density modification. The folds of individual subunits and sequence register at the initial steps of model building were determined using the positions of anomalous density peaks from Se atoms, directionality of helices and density connectivity. A combination of manual and automatic model building was used. Data collection and refinement statistics are presented in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 March; accepted 24 June 2011.

Published online 7 August 2011.

- Walker, J. E. The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. *Q. Rev. Biophys.* **25**, 253–324 (1992).
- Yagi, T. & Matsuno-Yagi, A. The proton-translocating NADH-quinone oxidoreductase in the respiratory chain: the secret unlocked. *Biochemistry* **42**, 2266–2274 (2003).
- Brandt, U. Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* **75**, 69–92 (2006).
- Ohnishi, T. Iron-sulfur clusters/semiquinones in complex I. *Biochim. Biophys. Acta* **1364**, 186–206 (1998).
- Sazanov, L. A. Respiratory complex I: mechanistic and structural insights provided by the crystal structure of the hydrophilic domain. *Biochemistry* **46**, 2275–2288 (2007).
- Schapiro, A. H. Human complex I defects in neurodegenerative diseases. *Biochim. Biophys. Acta* **1364**, 261–270 (1998).
- Dawson, T. M. & Dawson, V. L. Molecular pathways of neurodegeneration in Parkinson's disease. *Science* **302**, 819–822 (2003).
- Balaban, R. S., Nemoto, S. & Finkel, T. Mitochondria, oxidants, and aging. *Cell* **120**, 483–495 (2005).
- Carroll, J. *et al.* Bovine complex I is a complex of 45 different subunits. *J. Biol. Chem.* **281**, 32724–32727 (2006).
- Yip, C. Y., Harbour, M. E., Jayawardena, K., Fearnley, I. M. & Sazanov, L. A. Evolution of respiratory complex I: “supernumerary” subunits are present in the alpha-proteobacterial enzyme. *J. Biol. Chem.* **286**, 5023–5033 (2011).
- Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465**, 441–445 (2010).
- Sazanov, L. A. & Hinchliffe, P. Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* **311**, 1430–1436 (2006).
- Berrisford, J. M. & Sazanov, L. A. Structural basis for the mechanism of respiratory complex I. *J. Biol. Chem.* **284**, 29773–29783 (2009).
- Fearnley, I. M. & Walker, J. E. Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim. Biophys. Acta* **1140**, 105–134 (1992).
- Mathiesen, C. & Hagerhall, C. Transmembrane topology of the Nuol, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. *Biochim. Biophys. Acta* **1556**, 121–132 (2002).
- Friedrich, T. Complex I: a chimaera of a redox and conformation-driven proton pump? *J. Bioenerg. Biomembr.* **33**, 169–177 (2001).
- Hunte, C., Zickermann, V. & Brandt, U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* **329**, 448–451 (2010).
- Baranova, E. A., Holt, P. J. & Sazanov, L. A. Projection structure of the membrane domain of *Escherichia coli* respiratory complex I at 8 Å resolution. *J. Mol. Biol.* **366**, 140–154 (2007).
- von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487–494 (1992).
- Yau, W. M., Wimley, W. C., Gawrisch, K. & White, S. H. The preference of tryptophan for membrane interfaces. *Biochemistry* **37**, 14713–14718 (1998).
- Vik, S. B. The transmembrane helices of the L, M, and N subunits of Complex I from *E. coli* can be assigned on the basis of conservation and hydrophobic moment analysis. *FEBS Lett.* **585**, 1180–1184 (2011).
- Vinothkumar, K. R. & Henderson, R. Structures of membrane proteins. *Q. Rev. Biophys.* **43**, 65–158 (2010).
- Scrapanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261–267 (2007).
- Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
- Cooley, R. B., Arp, D. J. & Karplus, P. A. Evolutionary origin of a secondary structure: π -helices as cryptic but widespread insertional variations of α -helices that enhance protein functionality. *J. Mol. Biol.* **404**, 232–246 (2010).

26. Morino, M. *et al.* Single site mutations in the hetero-oligomeric Mrp antiporter from alkaliphilic *Bacillus pseudofirmus* DF4 that affect Na⁺/H⁺ antiport activity, sodium exclusion, individual Mrp protein levels, or Mrp complex formation. *J. Biol. Chem.* **285**, 30942–30950 (2010).
27. Nakamaru-Ogiso, E. *et al.* the membrane subunit NuoL (ND5) is involved in the indirect proton pumping mechanism of *Escherichia coli* complex I. *J. Biol. Chem.* **285**, 39070–39078 (2010).
28. Amarnah, B. & Vik, S. B. Mutagenesis of subunit N of the *Escherichia coli* complex I. Identification of the initiation codon and the sensitivity of mutants to decylubiquinone. *Biochemistry* **42**, 4800–4808 (2003).
29. Mathiesen, C. & Hagerhall, C. The 'antiporter module' of respiratory chain complex I includes the MrpC/NuoK subunit—a revision of the modular evolution scheme. *FEBS Lett.* **549**, 7–13 (2003).
30. Euro, L., Belevich, G., Verkhovskiy, M. I., Wikstrom, M. & Verkhovskaya, M. Conserved lysine residues of the membrane subunit NuoM are involved in energy conversion by the proton-pumping NADH:ubiquinone oxidoreductase (complex I). *Biochim. Biophys. Acta* **1777**, 1166–1172 (2008).
31. Torres-Bacete, J., Nakamaru-Ogiso, E., Matsuno-Yagi, A. & Yagi, T. Characterization of the NuoM (ND4) subunit in *Escherichia coli* NDH-1: conserved charged residues essential for energy-coupled activities. *J. Biol. Chem.* **282**, 36914–36922 (2007).
32. Kao, M. C., Nakamaru-Ogiso, E., Matsuno-Yagi, A. & Yagi, T. Characterization of the membrane domain subunit NuoK (ND4L) of the NADH-quinone oxidoreductase from *Escherichia coli*. *Biochemistry* **44**, 9545–9554 (2005).
33. Kervinen, M., Patsi, J., Finel, M. & Hassinen, I. E. A pair of membrane-embedded acidic residues in the NuoK subunit of *Escherichia coli* NDH-1, a counterpart of the ND4L subunit of the mitochondrial complex I, are required for high ubiquinone reductase activity. *Biochemistry* **43**, 773–781 (2004).
34. Fisher, N. & Rich, P. R. A motif for quinone binding sites in respiratory and photosynthetic systems. *J. Mol. Biol.* **296**, 1153–1162 (2000).
35. Ohnishi, T., Nakamaru-Ogiso, E. & Ohnishi, S. T. A new hypothesis on the simultaneous direct and indirect proton pump mechanisms in NADH-quinone oxidoreductase (complex I). *FEBS Lett.* **584**, 4131–4137 (2010).
36. Nakamaru-Ogiso, E., Sakamoto, K., Matsuno-Yagi, A., Miyoshi, H. & Yagi, T. The ND5 subunit was labeled by a photoaffinity analogue of fenpyroximate in bovine mitochondrial complex I. *Biochemistry* **42**, 746–754 (2003).
37. Steimle, S. *et al.* The role of subunit NuoL for proton translocation by the respiratory complex I. *Biochemistry* **50**, 3386–3393 (2011).
38. Michel, J., Deleon-Rangel, J., Zhu, S., Van Ree, K. & Vik, S. B. mutagenesis of the L, M, and N subunits of complex I from *Escherichia coli* indicates a common role in function. *PLoS ONE* **6**, e17420 (2011).
39. Krishnamoorthy, G. & Hinkle, P. C. Studies on the electron transfer pathway, topography of iron-sulfur centers, and site of coupling in NADH-Q oxidoreductase. *J. Biol. Chem.* **263**, 17566–17575 (1988).
40. Kao, M. C. *et al.* Characterization of the membrane domain subunit NuoJ (ND6) of the NADH-quinone oxidoreductase from *Escherichia coli* by chromosomal DNA manipulation. *Biochemistry* **44**, 3562–3571 (2005).
41. Patsi, J., Kervinen, M., Finel, M. & Hassinen, I. E. Leber hereditary optic neuropathy mutations in the ND6 subunit of mitochondrial complex I affect ubiquinone reduction kinetics in a bacterial model of the enzyme. *Biochem. J.* **409**, 129–137 (2008).
42. Galkin, A. S., Grivennikova, V. G. & Vinogradov, A. D. H⁺/2e⁻ stoichiometry in NADH-quinone reductase reactions catalyzed by bovine heart submitochondrial particles. *FEBS Lett.* **451**, 157–161 (1999).
43. Treberg, J. R. & Brand, M. D. A model of the proton translocation mechanism of complex I. *J. Biol. Chem.* **286**, 17579–17584 (2011).
44. Ohnishi, S. T., Salerno, J. C. & Ohnishi, T. Possible roles of two quinone molecules in direct and indirect proton pumps of bovine heart NADH-quinone oxidoreductase (complex I). *Biochim. Biophys. Acta* **1797**, 1891–1893 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Medical Research Council. We thank the ESRF and the SLS for provision of synchrotron radiation facilities. We are grateful to the staff of beamlines ID29 (ESRF) and X06SA (SLS) for assistance.

Author Contributions R.G.E. performed research and analysed data; L.A.S. designed the project, analysed data and wrote the manuscript, with contributions from R.G.E.

Author Information The coordinates and structure factors have been deposited in the RCSB Protein Data Bank under accession code 3RKO. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.A.S. (sazanov@mrcc-mbu.cam.ac.uk).

METHODS

Protein purification, crystallization and post-crystallization treatment. Complex I from *E. coli* was purified from strain BL21(DE3) as previously described⁴⁵. The membrane and hydrophilic domains were separated¹¹ by treatment with a high concentration of Mg^{2+} : purified complex I (2.5 mg ml^{-1}) was incubated in 20 mM Bis-Tris (pH 6.0), 5% (v/v) glycerol, 400 mM $MgCl_2$, 12.5 mM NaCl, 0.5 mM $CaCl_2$ and 0.25% (w/v) *n*-dodecyl- β -maltoide (Glycon) for 2 h on ice. The protein solution was diluted tenfold with buffer A (20 mM Bis-Tris (pH 6.0), 0.02% cymal-7 (Anatrace)), loaded onto a Mono-S HR 5/5 column and eluted with a 25 ml gradient, from 10% to 70%, of buffer B (1 M NaCl in buffer A). Fractions containing the membrane arm were pooled, concentrated in 100-kDa MWCO concentrators (Ultracel-100K, Amicon), diluted tenfold with 10 mM sodium acetate (pH 4.8), 50 mM NaCl and 0.02% cymal-7, and concentrated to $10\text{--}11\text{ mg ml}^{-1}$.

Before crystallization, lipids (1,2-dimyristoleoyl-*sn*-glycerol-3-phosphocholine with *E. coli* polar lipids at a 3:1 ratio (w/w), Avanti Polar Lipids) were added to the protein to 2 mg ml^{-1} final concentration. Crystals were grown by the sitting drop method, mixing protein at 1:1 (v/v) ratio with 0.1 M sodium acetate (pH 4.8), 0.3 M sodium tartrate (pH 5.0), 9% (w/v) PEG 4000, 0.15% sodium cholate and either 0.1 M NDSB-256 or 0.4% nonyl-glucoside from Detergent Screen HT (Hampton Research). Crystals grew at 22°C as thin plates of up to $200\text{ }\mu\text{m}$ in length and $20\text{--}40\text{ }\mu\text{m}$ in thickness. They were harvested after 3 weeks to 4 months, and initially diffracted to $7\text{ }\text{\AA}$ resolution. To improve diffraction properties, crystals were slowly cooled to 4°C , placed into micro-dialysis buttons, and mother liquor was exchanged by dialysis for 0.1 M sodium acetate (pH 4.8), 0.15 M sodium tartrate (pH 5.0), 30% PEG 4000, 0.1% sodium cholate and 12.5% glycerol over a period of 5 days.

For selenomethionine labelling, a feedback inhibition method⁴⁶ was applied during cell growth on M9-derived minimal media with 0.6% malate (optimized for maximal yield of complex I) as a carbon source. Purification and crystallization conditions were identical to those for native protein except for the addition of 10 mM tris(2-carboxyethyl)phosphine (TCEP, Thermo Scientific) to crystallization solutions.

Because the crystals were obtained at low pH and it is known that *E. coli* complex I dissociates into hydrophilic and hydrophobic domains after prolonged incubation at pH outside the 5.5–6.5 range⁴⁷, we assayed the purified enzyme for NADH:DQ activity at pH 4.8. The activity is lower than at optimal pH⁴⁵, but is still substantial ($\sim 4\text{ }\mu\text{mol}$ NADH oxidised per min per mg protein). It is still fully inhibited by piericidin A, indicating that enzyme is functional and that there are no drastic effects of pH on the conformation of the entire complex or individual domains.

Data collection and processing. Data were collected at 100 K with an ADSC Q210 detector at beamline ID29 at the European Synchrotron Radiation Facility and a Marmosaic 225 detector on the high-resolution diffractometer of beamline X06SA at the Swiss Light Source. Image data were processed with MOSFLM and SCALA from the CCP4 suite⁴⁸.

Model building and refinement. Crystals grew in space group *P1* with two molecules in the unit cell and 70% solvent content. As is the case with orthorhombic crystals¹¹, triclinic crystals were non-isomorphous and diffracted anisotropically, with the worst diffraction along the *a* axis and comparable resolution along axes *b* and *c*; the *c* axis being systematically slightly better. Using the Diffraction Anisotropy Server (<http://services.mbi.ucla.edu/anisotropy/>)⁴⁹, the best native data set was anisotropically scaled and truncated to $3.4\text{ }\text{\AA}$, $3.0\text{ }\text{\AA}$ and $3.0\text{ }\text{\AA}$ resolution, where the *F*/ σ ratio drops to $\sim 2.6\text{--}2.8$ along the *a**, *b** and *c** axes, respectively (scaling 2, Supplementary Table 1), resulting in better-quality electron density maps. As a control, the final model was also refined against non-truncated raw data at $3.2\text{ }\text{\AA}$ resolution, with excellent statistics (scaling 1, Supplementary Table 1). Crystal packing resembles that observed previously in space group *P2*₁*2*₁*2*₁, with strong interactions between hydrophilic surfaces and no visible side contacts, explaining the weaker diffraction along the *a* axis (Supplementary Fig. 2). It is possible that side contacts are mediated by poorly ordered lipids.

Several anomalous diffraction data collected from selenomethionine-containing crystals were used for phasing (Supplementary Table 1). Despite challenges presented by anomalous data collection in the *P1* space group, several good quality MAD data sets were obtained using a low-intensity beam and, in some cases, inverse-beam geometry. Initially, the positions of more than 140 Se atoms in the unit cell were determined and refined automatically in PHASER⁵⁰, using molecular replacement (MR) phases from the backbone model of the membrane fragment of *E. coli* complex I (PDB code 3M9C, ref. 11). The correspondence of the Se positions with peaks on the anomalous difference map was verified manually. When such maps were calculated with phases improved by density modification, the positions of 80 Se atoms per molecule were revealed. Experimental MAD phases calculated with these Se atoms in SHARP⁵¹ gave a figure of merit of 0.23 in the resolution range between $20\text{ }\text{\AA}$ and $3.5\text{ }\text{\AA}$, with the anomalous phasing power from peak, inflection and remote data sets being 0.608, 0.368 and 0.195, respectively. Isomorphous phasing power for inflection and remote data sets was 0.158 and 0.198, respectively.

Phases were improved and extended by multi-crystal averaging with histogram matching and solvent flattening in the CCP4-suite program DMmulti. The molecular masks and transformation matrices were obtained from MR solutions. Four data sets were used: two from two *P1* crystals (one native and one with Se MAD phases, Supplementary Table 1), and two phased data sets in the *P2*₁*2*₁*2*₁ space group from our previous study (Supplementary Table 1 in ref. 11, crystals Native1 with TaBr1 MAD phases and Native3 with TaBr3 SAD phases). Additionally, a solvent-flipped map was calculated in CNS⁵² using MAD data from a single SeMet-containing crystal. The maps were of excellent quality; they had better connectivity in DMmulti maps and, occasionally, better side-chain density in CNS maps, so both maps were used during model building.

Individual subunit folds and sequence register at the initial steps of model building were deduced on the basis of Se atom positions, directionality of helices determined using the 'find_helices_strands' procedure in PHENIX⁵³, and local density averaging in O (ref. 54), as well as density connectivity. An initial model was built by combining manual building in O and Coot v.0.6.1 (ref. 55) with automatic model building in Buccaneer v.1.5 (ref. 56). The model was rebuilt manually during cycles of refinement. The structure was refined in PHENIX. A subset of 2% of structure factors was omitted from refinement and used for cross-validation (calculation of R_{free}). Simulated annealing was used in the early stages; secondary structure and non-crystallographic symmetry restraints were used throughout the refinement and TLS (translation, libration and screw-rotation displacements parametrization) was used during the final refinement. For initial refinement, Se MAD experimental or blurred phases from multi-crystal averaging were used, whereas final refinement was only against native data. Several well ordered internal water molecules were modelled on the basis of their well defined electron density, presence of coordinating residues and low B-factor upon refinement. The final model includes 1,952 residues, one cymal-7 molecule, eight water molecules and nine fragments of lipid aliphatic chains. The structure was validated in PROCHECK⁵⁷ and MOLPROBITY⁵⁸ and was found to be of better-than-average quality for the resolution. As calculated with PROCHECK, 87.0/12.3/0.3/0.0% of residues are in preferable/allowed/generously allowed/disallowed Ramachandran regions, respectively.

Bioinformatics. Searches for structural analogs in the PDB were performed using the SSM server (www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver). In the case of NuoK and NuoA, as expected for small transmembrane proteins, similar folds could be found in the PDB, but without any significant sequence similarity. Structure-based multiple sequence alignments were performed in CLUSTALW v1.83 (ref. 59) with the profile alignment option. Homology models of the 14 conserved TM helices of Mrp subunits A and D from *Bacillus subtilis* and *Bacillus pseudofirmus* were built with MODELLER 9v7 (ref. 60) using structures of the *E. coli* subunits NuoL and NuoM, respectively, as templates. Sequences for modelling were aligned with CLUSTALW, using 30 sequences for NuoL or NuoM in addition to Mrp sequences and the structure-based profile alignment option. Figures were prepared in PyMol.

45. Sazanov, L. A., Carroll, J., Holt, P., Toime, L. & Fearnley, I. M. A role for native lipids in the stabilization and two-dimensional crystallization of the *Escherichia coli* NADH-ubiquinone oxidoreductase (Complex I). *J. Biol. Chem.* **278**, 19483–19491 (2003).
46. Van Duynne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
47. Leif, H., Sled, V. D., Ohnishi, T., Weiss, H. & Friedrich, T. Isolation and characterization of the proton-translocating NADH: ubiquinone oxidoreductase from *Escherichia coli*. *Eur. J. Biochem.* **230**, 538–548 (1995).
48. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
49. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
51. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
52. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
53. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
54. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
57. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

58. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
59. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
60. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **Ch. 5**, Unit 5.6 (2006).

Protection of repetitive DNA borders from self-induced meiotic instability

Gerben Vader^{1*}, Hannah G. Blitzblau^{1*}, Mihoko A. Tame¹, Jill E. Falk^{1†}, Lisa Curtin^{1,2} & Andreas Hochwagen^{1†}

DNA double strand breaks (DSBs) in repetitive sequences are a potent source of genomic instability, owing to the possibility of non-allelic homologous recombination (NAHR). Repetitive sequences are especially at risk during meiosis, when numerous programmed DSBs are introduced into the genome to initiate meiotic recombination¹. In the repetitive ribosomal DNA (rDNA) array of the budding yeast *Saccharomyces cerevisiae*, meiotic DSB formation is prevented in part through Sir2-dependent heterochromatin formation^{2,3}. Here we show that the edges of the rDNA array are exceptionally susceptible to meiotic DSBs, revealing an inherent heterogeneity in the rDNA array. We find that this localized DSB susceptibility necessitates a border-specific protection system consisting of the meiotic ATPase Pch2 and the origin recognition complex subunit Orc1. Upon disruption of these factors, DSB formation and recombination increased specifically in the outermost rDNA repeats, leading to NAHR and rDNA instability. Notably, the Sir2-dependent heterochromatin of the rDNA itself was responsible for the induction of DSBs at the rDNA borders in *pch2Δ* cells. Thus, although the activity of Sir2 globally prevents meiotic DSBs in the rDNA, it creates a highly permissive environment for DSB formation at the junctions between heterochromatin and euchromatin. Heterochromatinized repetitive DNA arrays are abundant in most eukaryotic genomes. Our data define the borders of such chromatin domains as distinct high-risk regions for meiotic NAHR, the protection of which may be a universal requirement to prevent meiotic genome rearrangements that are associated with genomic diseases and birth defects.

To understand better the mechanisms that protect repetitive DNA from meiotic NAHR, we analysed the single tandem rDNA array of budding yeast. Meiotic DSB formation and recombination in the rDNA are repressed by the histone deacetylase Sir2 (refs 2, 3). Additionally, Pch2, a widely conserved meiosis-specific ATPase, suppresses meiotic recombination in the rDNA by an unknown mechanism^{4,5}. We used clamped-homogenous electric field (CHEF) electrophoresis and Southern blotting of excised rDNA arrays to address whether Pch2 regulates meiotic DSB formation in the rDNA. Consistent with previous results^{2,3}, the level of full-length rDNA arrays that remained 8 h after meiotic induction was significantly reduced in *sir2Δ* mutants compared to wild-type cells, indicating increased DSB formation (Fig. 1a and Supplementary Fig. 1a). By contrast, no such reduction occurred in *pch2Δ* mutants, although we observed a tenfold increase in crossover recombination across the rDNA array (Fig. 1a, b). Because small changes in array length would not be detectable by the CHEF gel assay, we wondered whether DSB formation in *pch2Δ* mutants occurred specifically in the outermost rDNA repeats. To test this possibility, we generated *pch2Δ* strains carrying a *URA3* insertion at defined positions in the rDNA array (Fig. 1c) and analysed the rDNA repeat units that directly flank these insertions by Southern blotting. We observed a strongly DSB-prone site in repeat 1 and weak DSB

formation in repeat 3, whereas no DSB formation was detectable in repeat 10 of the approximately 100 rDNA repeats (Fig. 1d). Thus, *pch2Δ* cells undergo increased meiotic DSB formation predominantly in the outermost rDNA repeats.

To determine whether *PCH2* suppresses DSB formation only within the rDNA, or also in other regions of the genome, we first analysed a chromosomal fragment spanning the junction between single-copy DNA and rDNA in a *pch2Δ* mutant by Southern blotting. We observed additional, strong DSB formation in the adjacent single-copy sequences (Fig. 1e and Supplementary Fig. 1b), which were previously shown to have exceptionally low levels of meiotic DSBs in *PCH2* cells^{6,7} (Fig. 1f). The observed break sites behaved similarly to known meiotic DSBs⁸; they were induced during meiosis in *dmc1Δ* and *DMC1* cells (Fig. 1d, e and Supplementary Fig. 1c), depended on the meiotic DSB machinery (Supplementary Fig. 1d)⁹, promoted meiotic recombination (Supplementary Fig. 1e) and occurred in gene promoters (Fig. 1e and Supplementary Fig. 1b). Indeed, even the DSBs observed in repeat 1 mapped to the promoter of a gene (*TAR1*) that is encoded in every rDNA repeat¹⁰ (Fig. 1e). Genome-wide analysis of DSBs⁶ in *pch2Δ* cells showed that strong DSB induction occurred in 30–50-kilobase (kb) regions of single-copy sequence abutting both sides of the rDNA (Fig. 1f). Increased DSB formation was also observed close to other heterochromatic regions (telomeres and *HML*), whereas the DSB landscape elsewhere in the genome was not markedly altered (Supplementary Figs 1f, g, 2 and Supplementary Table 1). In contrast to *pch2Δ* mutants, the loss of *SIR2* did not lead to increased DSB formation adjacent to the rDNA array (Fig. 1f). Thus, Pch2 represses recombination in the rDNA at the level of DSB formation, but in a manner distinct from Sir2.

We investigated whether the increased DSB formation in the outermost rDNA repeats in *pch2Δ* mutants (Fig. 1d) resulted in a local increase in rDNA recombination. We measured recombination rates using flanking markers to the left and right of the rDNA together with a collection of single *URA3* insertions tiling inwards from the left side of the rDNA (Fig. 1c). Analysis of a *URA3* insertion in the centre of the rDNA (inserted next to repeat 49 of 99) indicated that recombination occurred in a symmetrical pattern. Notably, about 80% of the recombination events in the left half of the rDNA occurred within the first ten repeats from the left border (Fig. 1g and Supplementary Table 2), with about 30% taking place within repeat 1. Thus, there is a strong bias for recombination in the rDNA repeats very close to the array border.

Because recombination in repetitive DNA can lead to NAHR, we selected tetrads of *pch2Δ* mutants that had undergone recombination in the rDNA, and determined the resulting rDNA repeat number between the *URA3* insertion and the left rDNA boundary. In 70% ($n = 47$) of the tetrads investigated from different *URA3* integrants, we detected changes in repeat number, ranging from 1 to 19 repeats (Fig. 1i, j and Supplementary Table 2), demonstrating that rDNA crossovers in *pch2Δ* cells are frequently associated with NAHR.

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Somerville High School, Somerville, Massachusetts 02143, USA. [†]Present addresses: David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (J.E.F.); Department of Biology, New York University, 100 Washington Square East, New York, New York 10003, USA (A.H.).

*These authors contributed equally to this work.

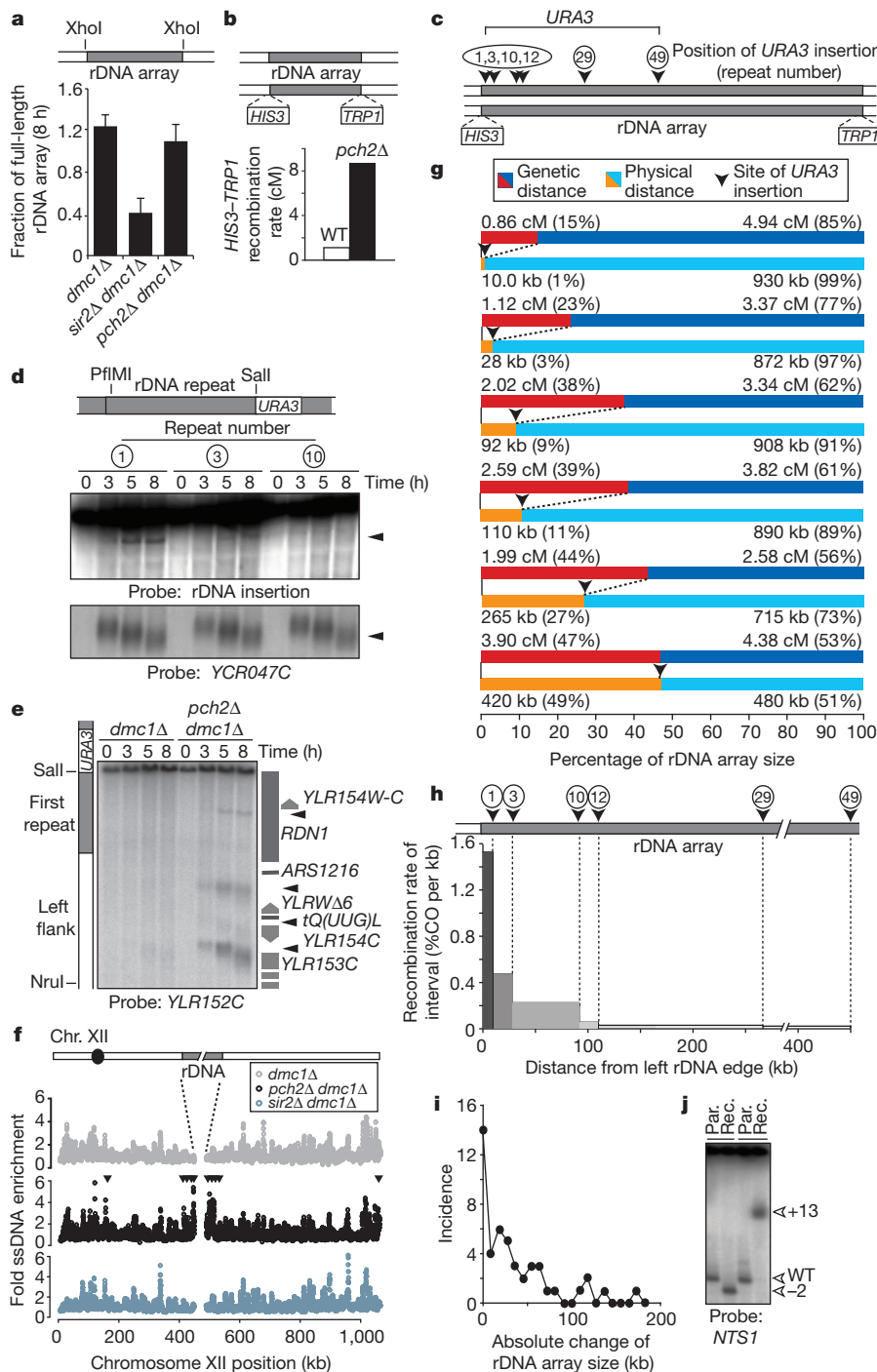


Figure 1 | Ribosomal-DNA-associated DSB formation and recombination.

a, CHEF analysis of the rDNA of meiotic *dmc1Δ* (H5217), *sir2Δ dmc1Δ* (H2953) and *pch2Δ dmc1Δ* (H5216) cells. The schematic shows the analysed XhoI restriction fragment. A *dmc1Δ* mutation was used to prevent DSB repair. The mean (+ s.e.m.) of five experiments is shown. Significance was assessed by one-tailed Student's *t*-test: *dmc1Δ* versus *sir2Δ dmc1Δ*, *P* value = 0.00122; *dmc1Δ* versus *pch2Δ dmc1Δ*, *P* value = 0.254; *pch2Δ dmc1Δ* versus *sir2Δ dmc1Δ*, *P* value = 0.00216. **b**, Schematic of markers inserted in unique single-copy sequences within 500 base pairs (bp) of the rDNA, and crossover rates in wild-type (WT; H3026; *n* = 467) and *pch2Δ* (H3027; *n* = 186) cells. **c**, Schematic indicating marker locations in the rDNA used in **d** and **g**. *URA3* markers were inserted in the *NTS1/2* region of the indicated repeats. **d**, Southern blot for restriction fragments containing the indicated insert-associated rDNA repeat units from *pch2Δ dmc1Δ* strains H5622 (repeat 1), H5636 (repeat 3) and H5706 (repeat 10). DNA was digested with PflMI and SalI and probed for the unique rDNA insertion. The *YCR047C* probe (on DNA digested with HindIII) was a positive control for DSB formation. **e**, Southern blot of the left rDNA flank,

including the outermost rDNA repeat, in *dmc1Δ* (H5583) and *pch2Δ dmc1Δ* (H5622) cells (SalI and NruI digest; probe *YLR152C*). Positions of open reading frames are shown schematically alongside the Southern blot. **f**, Enrichment profile of ssDNA in chromosome XII in *dmc1Δ* (H118, light grey), *pch2Δ dmc1Δ* (H2629, black) and *sir2Δ dmc1Δ* (H2953, dark grey) cells. Arrowheads indicate >twofold increased DSB formation in *pch2Δ dmc1Δ* compared to *dmc1Δ* cells. **g**, Tetrad analysis of *pch2Δ URA3*-rDNA-insertion strains H4611 (repeat 1), H4613 (repeat 3), H3823 (repeat 10), H4612 (repeat 12), H3820 (repeat 29) and H3821 (repeat 49; Supplementary Table 2). Recombination rates between *URA3* and rDNA-flanking markers are shown in relation to the physical *URA3* positions in the rDNA. **h**, Relative contribution of each measured interval indicated in **g** to total rDNA recombination (percentage of crossovers (CO) per kb of interval). **i**, Incidence of changes in rDNA repeat number between the *URA3* insertion and the rDNA boundary in *pch2Δ* tetrads that had undergone crossover recombination. **j**, CHEF analysis of two tetrads that have undergone unequal recombination with parental controls. DNA was digested with XhoI and probed with *NTS1*. Par., parental controls; Rec., recombinants.

However, the prevalence of allelic recombinants (30%) and the fact that changes in repeat number encompassed less than 20% of the approximately 100–110 rDNA repeats in our strains indicate that the homology search for DSB repair in the outermost rDNA repeats in *pch2Δ* cells is restricted to close neighbours. Notably, the distribution of changes in repeat number closely matched the pattern of crossover events (Fig. 1h, i). This congruence indicates that the spread in the crossover distribution (Fig. 1g, h) can largely be accounted for by non-allelic exchanges between rDNA repeats originating from DSBs in the outermost repeats of the array. Finally, although rDNA exchanges occurred at a much lower frequency in wild-type cells, they were also associated with NAHR (Supplementary Table 2), indicating that Pch2 primarily acts to prevent NAHR by suppressing DSB formation. These results establish the rDNA borders as high-risk regions for meiotic NAHR.

To determine how Pch2 suppresses DSB formation near the rDNA, we measured the chromosome association of DSB-related factors at the time of DSB formation. The three essential DSB factors^{11,12} that we were able to analyse by chromatin immunoprecipitation (ChIP), Rec114, Mer2 and Mre11, were specifically enriched near the rDNA and *HML* in *pch2Δ* cells, mirroring the changes in DSB formation (Fig. 2 and Supplementary Fig. 3). We then investigated whether the regional exclusion of DSB factors could be explained by local depletion of the DSB-promoting chromosome-axis protein Hop1, the cytological distribution of which is affected by Pch2 (refs 4, 13). Although Hop1 binding was slightly increased near the rDNA and *HML* in *pch2Δ* cells, it was abundant even in wild-type cells, indicating that Pch2 does not regulate the initial chromosomal recruitment of Hop1. Rather, the differences in Hop1 binding that we observed might reflect an effect of Pch2 on chromosome structure¹³. Finally, although DSBs are enriched in promoters containing histone H3 lysine 4 trimethylation (H3K4me3)¹⁴, we saw no difference in the genome-wide levels of this modification with or without Pch2 (Fig. 2c and Supplementary Figs 3d and 4), indicating that Pch2 does not influence this chromatin modification. These findings indicate that Pch2 specifically blocks the stable recruitment of DSB factors to prevent local DSB formation.

We sought to identify proteins that collaborate with Pch2 in preventing rDNA-proximal DSBs. A yeast two-hybrid screen isolated a fragment of the Orc1 protein, containing its ATPase domain, as a Pch2 interactor (Fig. 3a). This interaction was confirmed by co-immunoprecipitation (Fig. 3b and Supplementary Fig. 5a). Orc1 is a component of the conserved origin recognition complex that has several important chromosomal functions, including the loading of the replicative helicase¹⁵. Impairing Orc1 protein levels by a temperature-sensitive *orc1-161* mutation¹⁶ (Supplementary Fig. 5b) triggered DSB formation in the rDNA flanking regions, similarly to the loss of *PCH2* (Fig. 3c, d). DSB formation near the rDNA occurred even at a temperature (23 °C) that is permissive for pre-meiotic DNA replication and spore viability (Fig. 3c–e and Supplementary Fig. 5c, d). Similarly, we saw increased DSB levels near the rDNA in an *orc1* mutant lacking the N-terminal bromo-adjacent homology (BAH) domain that is required for the chromatin-silencing function of Orc1 (ref. 17), but is dispensable for DNA replication (Fig. 3f and Supplementary Fig. 5e, f). These data indicate that the regulatory roles of Orc1 in DSB formation and bulk DNA replication are separable, although we cannot rule out that the analysed *orc1* mutations affect rDNA replication locally. During meiosis, Pch2 concentrates in the nucleolus, the organelle assembled on the rDNA array⁴. In *orc1-161* cells, the recruitment of Pch2 to the nucleolus was impaired, despite normal levels of cellular Pch2 (Fig. 3g, h). Both Pch2 and Orc1 belong to the AAA⁺ family of ATPases that often function as multimeric complexes¹⁸, and we found that the ATPase activity of Pch2 was required to prevent rDNA-proximal DSBs (Supplementary Fig. 5g). These data define a role for Orc1 in the nucleolar recruitment and possible activation of Pch2 to prevent local DSB formation.

To find out whether the specific DSB activity at the rDNA borders in *pch2Δ* mutants was linked to the presence of the rDNA itself, we

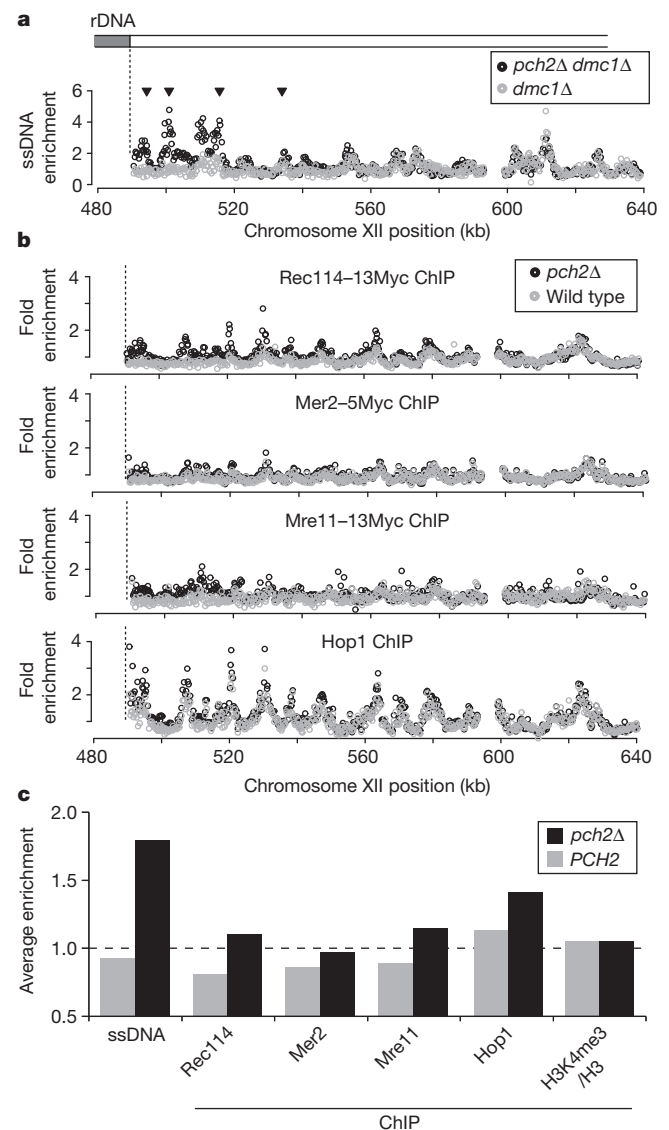


Figure 2 | Association of the meiotic DSB machinery near the rDNA.

a, Enrichment of ssDNA in a region flanking the right border of the rDNA on chromosome XII (*Saccharomyces* genome database (SGD) coordinates) in *dmc1Δ* (H118, grey) and *pch2Δ dmc1Δ* (H2629, black) cells. Arrowheads indicate >twofold increased DSB formation in *pch2Δ dmc1Δ* compared to *dmc1Δ* cells. **b**, ChIP-chip analysis in same region as in **a**, for the proteins Rec114–13Myc (first panel; wild type (H4890), *pch2Δ* (H4893)), Mer2–5Myc (second panel; wild type (H5917), *pch2Δ* (H5916)), Mre11–13Myc (third panel; wild type (H5547), *pch2Δ* (H5947)) and Hop1 (fourth panel; wild type (H119), *pch2Δ* (H2817)), in wild-type (grey) and *pch2Δ* (black) cells. **c**, Average enrichment for the different *PCH2* and *pch2Δ* ssDNA and ChIP data sets (see **a**, **b** and Supplementary Fig. 4) within the 50 kb flanking the right rDNA border (see Methods for genomic coordinates). The genome-wide average is indicated by the dotted line.

deleted the rDNA array from its genomic location. In these strains, loss of *PCH2* no longer allowed DSB formation in the flanking regions (Figs 4a, b), demonstrating an intrinsic DSB-promoting activity in the rDNA. To investigate whether the rDNA was sufficient to promote DSB formation, we created a translocation between chromosomes II and XII that exchanged the rDNA-distal portion of chromosome XII with a portion of chromosome II (Fig. 4c). In these strains, DSB levels were no longer increased in the former right flank of the rDNA (now flanked by chromosome II sequences; Fig. 4d), but notably, increased DSB formation was observed on the chromosome II sequences that, after translocation, were flanking the rDNA (Fig. 4e). Thus, in the absence of *PCH2*, the rDNA is necessary and sufficient to promote DSB formation.

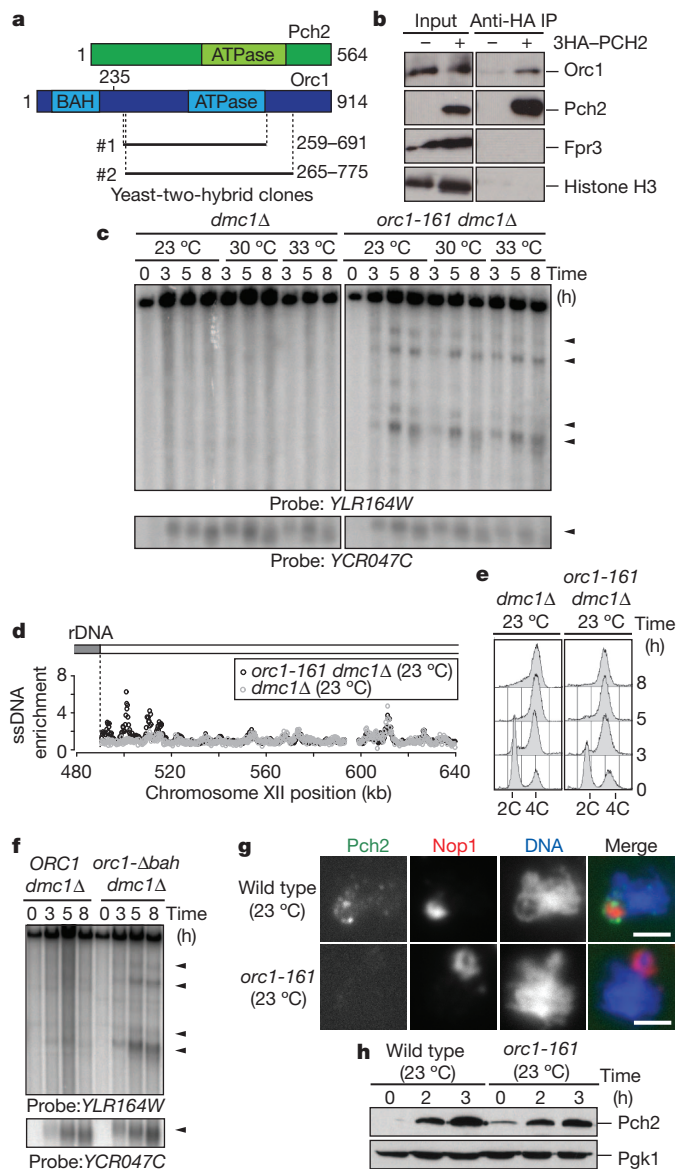


Figure 3 | Orc1 and Pch2 collaborate to suppress DSB formation.

a, Schematic of Pch2 and Orc1 proteins, indicating Orc1 clones identified by yeast two-hybrid screen. **b**, Co-immunoprecipitation (IP) between haemagglutinin-tagged Pch2 (3HA-Pch2) and Orc1 in wild-type (H119) and 3HA-PCH2 (H3463) cells. Fpr3 and histone H3 are controls for the nucleolar and chromosomal fractions, respectively. **c**, Southern blots of the right rDNA flank and YCR047C in meiotic *dmc1Δ* (H118) and *orc1-161 dmc1Δ* (H4952) cells, grown at the indicated temperatures. Arrowheads indicate broken DNA fragments. **d**, Profiles of ssDNA in the region flanking the right rDNA border in *orc1-161 dmc1Δ* (H5137, black) and *dmc1Δ* (H118, grey) cells, grown at 23 °C. **e**, DNA-content analysis of meiotic *dmc1Δ* (H118) and *orc1-161 dmc1Δ* (H5137) cells grown at 23 °C. 2C and 4C refer to unreplicated and replicated diploid DNA contents, respectively. **f**, Southern blots of the right rDNA flank and YCR047C in *ORC1 dmc1Δ* (H5838) and *orc1-Δbah dmc1Δ* (H5865) cells. **g**, Immunofluorescence of chromosome spreads stained for Pch2 (HA, green), Nop1 (nucleolar marker, red) and DNA (blue) in 3HA-PCH2 (H3463) and 3HA-PCH2 *orc1-161* (H5033) cells at 3 h after meiotic induction at 23 °C. Scale bar, 2 μm. **h**, Western blot analysis showing Pch2 (HA) expression in 3HA-PCH2 (H3463) and 3HA-PCH2 *orc1-161* (H5033) cells. Pgk1 is used as loading control.

The rDNA is assembled into specialized, Sir2-dependent chromatin, and we asked how this chromatin state influenced DSB formation at the rDNA boundaries. Notably, loss of Sir2 protein or loss of its deacetylase activity¹⁹ largely eliminated DSB formation in the rDNA flanking regions in *pch2Δ* mutants (Fig. 4f, g). Thus, although

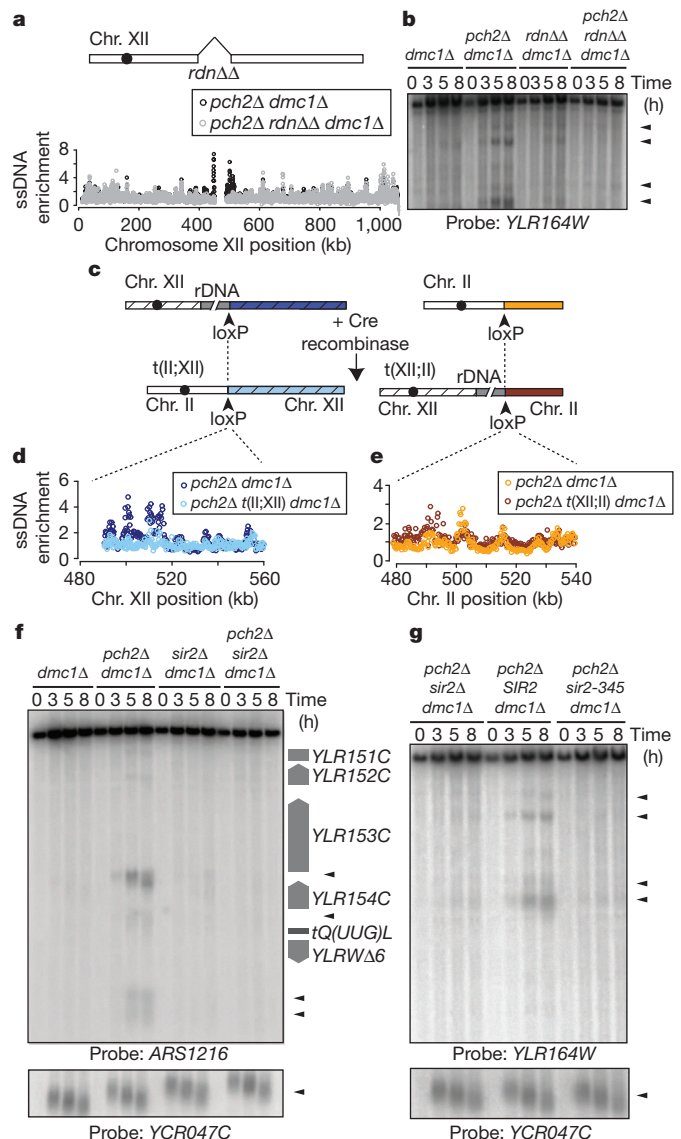


Figure 4 | Ribosomal-DNA chromatin promotes DSB formation.

a, Schematic of the rDNA deletion strain (*rdnΔΔ*) and ssDNA profiles of the region flanking the right rDNA border in *pch2Δ dmc1Δ* (H2629, black) and *pch2Δ rdnΔΔ dmc1Δ* (H4737, grey) cells. **b**, Southern blot analysis of the right rDNA flank in *dmc1Δ* (H118), *pch2Δ dmc1Δ* (H4736) and *pch2Δ rdnΔΔ dmc1Δ* (H4737) cells. Arrowheads indicate broken DNA fragments. **c**, Strategy used to generate chromosomal translocations between chromosomes XII and II. **d**, **e**, Profiles of ssDNA in strains containing the XII;II translocation. In **d**, the depicted region is next to the rDNA in *pch2Δ dmc1Δ* (H2629) cells (dark blue) and next to the left arm of chromosome II in *pch2Δ t(II;XII) dmc1Δ* (H4798) cells (light blue). In **e**, the depicted region is located on chromosome II in *pch2Δ dmc1Δ* cells (orange) and next to rDNA in *pch2Δ t(XII;II) dmc1Δ* cells (dark red). **f**, Southern blot of the left rDNA flank (HindIII digest; ARS1216 probe) and of YCR047C in *dmc1Δ* (H118), *pch2Δ dmc1Δ* (H2629), *sir2Δ dmc1Δ* (H2953) and *pch2Δ sir2Δ dmc1Δ* (H3038) cells. Positions of open reading frames are shown schematically alongside the Southern blot. **g**, Southern blot of the right rDNA flank and YCR047C in *pch2Δ sir2Δ dmc1Δ* (H3262), *pch2Δ sir2Δ dmc1Δ leu2::SIR2* (H3261) and *pch2Δ sir2Δ dmc1Δ leu2::sir2-345* (H3282) cells. Arrowheads indicate broken DNA fragments.

Sir2-dependent heterochromatin suppresses meiotic DSBs within the rDNA array (refs 2, 3 and Fig. 1a), it has a profound DSB-promoting effect on the rDNA borders that is counteracted by Pch2 and Orc1. It is also notable that Sir2 itself localizes Pch2 to the nucleolus⁴, reflecting an elegant coupling mechanism that maintains meiotic stability across the entire rDNA. The double dependence of Pch2 on Sir2 and Orc1

may promote Pch2 enrichment at the nucleolus, analogous to the bimodal recruitment mechanisms that restrict localization of Aurora B and shugoshin to centromeres²⁰.

Although repeat-associated chromatin marks differ substantially between organisms and even among individual loci²¹, the assembly of heterochromatin on repetitive DNA arrays is a common strategy to protect the genome against destabilization caused by errors in meiotic recombination¹. Our results establish borders between heterochromatin and euchromatin as potential high-risk regions for meiotic DSB formation and NAHR, and reveal the existence of a secondary border-specific system that shields against these events. Buffer zones like those established by Pch2 and Orc1 may need to be broad, because even DSBs adjacent to repetitive DNA can trigger NAHR²². Given the prominent presence of repetitive DNA arrays in genomes ranging from yeast to man¹, we propose that mechanisms that limit DSB activity around repetitive DNA might be a widespread phenomenon.

METHODS SUMMARY

Yeast strains were of the SK1 background and are listed in Supplementary Table 3. Analysis of single-stranded DNA profiles and ChIP-chip analysis were performed as previously described^{6,23}. These and other standard techniques used are detailed in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 January; accepted 24 June 2011.

Published online 7 August 2011.

1. Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nature Rev. Mol. Cell Biol.* **11**, 182–195 (2010).
2. Gottlieb, S. & Esposito, R. E. A new role for a yeast transcriptional silencer gene, *SIR2*, in regulation of recombination in ribosomal DNA. *Cell* **56**, 771–776 (1989).
3. Mieczkowski, P. A. *et al.* Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **104**, 3955–3960 (2007).
4. San-Segundo, P. A. & Roeder, G. S. Pch2 links chromatin silencing to meiotic checkpoint control. *Cell* **97**, 313–324 (1999).
5. Wu, H. Y. & Burgess, S. M. Two distinct surveillance mechanisms monitor meiotic chromosome metabolism in budding yeast. *Curr. Biol.* **16**, 2473–2479 (2006).
6. Blitzblau, H. G. *et al.* Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr. Biol.* **17**, 2003–2012 (2007).
7. Gerton, J. L. *et al.* Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **97**, 11383–11390 (2000).
8. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
9. Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
10. Coelho, P. S. *et al.* A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA. *Genes Dev.* **16**, 2755–2760 (2002).
11. Arora, C., Kee, K., Maleki, S. & Keeney, S. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol. Cell* **13**, 549–559 (2004).
12. Keeney, S. & Neale, M. J. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem. Soc. Trans.* **34**, 523–525 (2006).
13. Borner, G. V., Barot, A. & Kleckner, N. Yeast Pch2 promotes domainal axis organization, timely recombination progression, and arrest of defective recombinosomes during meiosis. *Proc. Natl Acad. Sci. USA* **105**, 3327–3332 (2008).
14. Borde, V. *et al.* Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* **28**, 99–111 (2009).
15. Bell, S. P. The origin recognition complex: from simple origins to complex functions. *Genes Dev.* **16**, 659–672 (2002).
16. Gibson, D. G., Bell, S. P. & Aparicio, O. M. Cell cycle execution point analysis of ORC function and characterization of the checkpoint response to ORC inactivation in *Saccharomyces cerevisiae*. *Genes Cells* **11**, 557–573 (2006).
17. Bell, S. P. *et al.* The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83**, 563–568 (1995).
18. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
19. Imai, S., Armstrong, C. M., Kaerberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
20. Vader, G. & Lens, S. M. Chromosome segregation: taking the passenger seat. *Curr. Biol.* **20**, R879–R881 (2010).
21. Moazed, D. Common themes in mechanisms of gene silencing. *Mol. Cell* **8**, 489–498 (2001).
22. Hoang, M. L. *et al.* Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. *PLoS Genet.* **6**, e1001228 (2010).
23. Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59–69 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. P. Bell, A. Shinohara, N. Hunter, N. Hollingsworth and F. Klein for sharing reagents and data. We thank I. Cheeseman, M. Gehring and V. Subramanian for discussions and critical reading of the manuscript. This work was supported by NIH grant GM088248 to A.H. and by fellowships from the Netherlands Organisation for Scientific Research (NWO Rubicon-825.08.009 and NWO VENI-016.111.004) to G.V.; L.C. was supported by an HHMI Institutional Undergraduate Education Grant to MIT (grant 52005879).

Author Contributions G.V., H.G.B. and A.H. designed and performed experiments and analysed the data. M.A.T. performed the yeast two-hybrid analysis. J.E.F., L.C. and A.H. performed recombination mapping. G.V., H.G.B. and A.H. wrote the paper.

Author Information All data sets in this publication are available in the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>), accession number GSE30073. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.H. (andi@nyu.edu).

METHODS

Yeast strains and two-hybrid analysis. All yeast strains used in this study were constructed in the SK1 background and are listed in Supplementary Table 3. Epitope tags and gene disruptions were introduced by standard PCR-based transformation. The *orc1-Abah* mutant was generated using a plasmid encoding a truncated version of Orc1 (amino acids 235–914; pSPB1.48, gift from S. P. Bell²⁴). To create *URA3* insertions in the rDNA, cells were transformed with a pRS306-NTS1/2 plasmid linearized with SphI (this plasmid contains a 2341-bp fragment harbouring the intergenic rDNA sequences, *NTS1* and *NTS2*, ligated into the BamHI and EcoRI sites of pRS306). Insertion sites in the rDNA were mapped by CHEF gel analysis using a unique XhoI site in the inserted sequence (XhoI does not cut in the rDNA), and suitable clones were selected for further analysis. SK1 strains lacking the rDNA array were generated as described in ref. 25. Briefly, cells were transformed with a very high-copy rDNA plasmid (*pRDN-hyg::URA3::leu2-8*) carrying a recessive point mutation that confers resistance to hygromycin²⁶. After selection on hygromycin, a clone was selected that had lost all but three repeats of the rDNA array through spontaneous deletion, as determined by CHEF gel analysis. The remaining rDNA copies were subsequently deleted by conventional gene disruption using a *HIS3* deletion cassette. Complete deletion of the rDNA array was confirmed by Southern blotting. Chromosomal translocations between chromosomes XII and II were generated essentially as previously described²⁷. Briefly, plasmids containing a promoter-less *ADE2* gene adjacent to a *loxP* site (*loxP-ADE2::natMX4*) and a *GPD* promoter (*pGPD*) with an adjacent *loxP* site (*pGPD-loxP::hphMX4*; both plasmids were gifts from N. Hunter) were integrated at *YLR162W-A* and *LYS2*, respectively. After induction of Cre recombinase from a pGAL-Cre plasmid (N. Hunter), cells were selected that had undergone translocation between *LYS2* (chromosome II) and *YLR162W-A* (chromosome XII). Translocation was confirmed by Southern blot analysis. To identify interactors of Pch2 by two-hybrid screen, full-length *PCH2* was amplified from genomic DNA and the intron was removed by site-directed mutagenesis. The Pch2 coding sequence was then cloned into pGBDU-C1, and the resulting bait plasmid was used to screen libraries in all three reading frames²⁸.

Synchronous meiosis. Cells were grown for 24 h in yeast peptone dextrose (YPD) at 23 °C then diluted in BYTA medium (1% yeast extract, 2% tryptone, 1% potassium acetate, 50 mM potassium phthalate) to a optical density at 600 nm (OD_{600}) of 0.3 ($OD_{600} = 0.5$ for *orc1-161*, *rdnΔΔ* and XII;II translocation strains), and grown for 16 h at 30 °C (or for 18 h at 23 °C in the case of temperature-sensitive strains). After two washes in water, cells were diluted into SPO medium (0.3% potassium acetate) at $OD_{600} = 1.9$ and incubated at 30 °C unless otherwise stated.

Isolation of ssDNA. For ssDNA analysis^{29,30}, about 10^9 cells were fixed in 70% ethanol at –20 °C at 0 h and 5 h after induction of meiosis. After spheroplasting in sorbitol buffer (1 M sorbitol, 1% β-mercaptoethanol, 0.2 mg ml^{–1} zymolyase, and 0.1 M EDTA, pH 7.4), cells were lysed in NDS buffer (0.6% SDS, 300 mM EDTA, 10 mM Tris-HCl, pH 9.5). After treatment with proteinase K (0.25 mg ml^{–1}) and RNase A, the DNA was digested with EcoRI and ssDNA was then enriched by adsorption to BND-cellulose and eluted using 1.8% caffeine. This enriched ssDNA was subsequently used for microarray analysis. For this, 1.5 μg of the respective 0-h and 5-h ssDNA samples was labelled with Cy3-dUTP or Cy5-dUTP (GE Healthcare) by random priming without denaturation using 4 μg of random primer (Integrated DNA Technologies) and 10 units of Klenow (New England Biolabs).

Western blotting and immunoprecipitation. For western blotting, 5 ml of meiotic cells were harvested at the indicated time points and resuspended in 5% trichloroacetic acid. After incubation on ice for 10 min, samples were washed in acetone and dried overnight. Samples were lysed by bead beating in a FastPrep FP120 (Thermo Scientific) in TE lysis buffer (10 mM Tris (pH 7.5), 1 mM EDTA, 2.75 mM dithiothreitol). SDS loading buffer (3×) was added and the pH of the sample was adjusted to neutral by addition of 1 M Tris (pH 8.0). Samples were separated by standard polyacrylamide gel electrophoresis. The following antibodies were used: anti-ORC (1108, 1:1,000, gift from S. P. Bell), anti-Fpr3 (1:1,000, gift from J. Thorner), anti-HA (3F10, 1:1,000, Roche), anti-Pgk1 (1:1,000, Invitrogen) and anti-histone H3 (1:1,000, Abcam). For immunoprecipitations, 50 ml of meiotic cells were harvested 3 h after induction of meiosis. Cells were diluted in 2× lysis buffer (20 mM HEPES (pH 7.5), 4 mM MgCl₂, 0.6 M glutamic acid, 0.32 M sorbitol, 4% glycerol, 0.5% Triton X-100) containing protease inhibitors, and lysed by bead beating. Extracts were sonicated and cleared by centrifugation. After removal of one tenth of the extract for an input sample, extracts were immunoprecipitated with 2 μl anti-HA (3F10, Roche) for 3 hA–Pch2, 2 μl anti-ORC (1108) and 2 μl anti-Fpr3, in combination with 20 μl of a 50% slurry of GammaBind-Sepharose beads (GE Healthcare) for 16 h at 4 °C. After five washes in 1× lysis buffer, 1× SDS loading buffer was added to the beads, and samples were analysed by western blotting with the indicated antibodies.

Chromatin immunoprecipitation. Meiotic cells (25 ml) were harvested 3 h after induction of meiosis and fixed for 15 min in 1% formaldehyde. The formaldehyde

was quenched by addition of 125 mM glycine. Samples were processed as previously described²³. Before immunoprecipitation, one tenth of the sample was removed as input sample. The antibodies used for immunoprecipitation were: 2 μl anti-Myc (9E11, Abcam; for Rec114–13Myc, Mer2–5Myc and Mre11–13Myc), 2 μl anti-Hop1 (gift from N. Hollingsworth), 2 μl anti-histone-H3 (AB1791, Abcam) and 2 μl anti-H3K4me3 (AB8580, Abcam), in combination with 20 μl of a 50% slurry of GammaBind-Sepharose beads (GE Healthcare). Half of the ChIP sample and one tenth of the input were labelled with Cy3-dUTP or Cy5-dUTP (GE Healthcare) as described in the ssDNA protocol, with the difference that the DNA was denatured for 5 min at 95 °C before the extension reaction.

Microarray analysis. After removal of unincorporated dyes, Cy3- and Cy5-labelled samples were hybridized to custom 4 × 44K tiled genomic yeast microarrays (Agilent Technologies) for 16 h at 65 °C. Levels of Cy3 and Cy5 were calculated with the Agilent Feature Extractor CGH software. Background normalization, log₂ ratios for each experiment and scale normalizations between experiments were calculated with the sma package in R (v2.1.0, <http://www.r-project.org>)^{29,30}. Each data set is an average of two experiments. For comparison between isogenic wild-type and *pch2Δ* cells, data sets were scale-normalized. To analyse the distribution of H3K4me3, we normalized the enrichment to that of total histone H3 generated from the same extracts, by subtracting the log₂ ratios. To measure the average ssDNA or ChIP enrichment in different chromosomal regions, the following SGD coordinates were analysed, on the basis of the positions of available array features:

50 kb right of the rDNA: XII, 490,531–540,530.

50 kb left of the rDNA: XII, 401,371–451,370.

Rest of chromosome XII: XII, 1–401,370 and 540,531–1,078,177.

First 100 kb of chromosome III: III, 1–100,000.

Rest of chromosome III: III, 100,001–316,620.

Chromosome VIII: VIII, 1–562,643.

Chromosome spreads and immunofluorescence. Meiotic cells were spread as described previously³¹. Cells were spheroplasted at 37 °C in solution 1 (2% potassium acetate, 0.8% sorbitol, 10 mM dithiothreitol, 130 mg ml^{–1} zymolyase 100T (Seikagaku)). Solution 2 (100 mM MES (pH 6.4), 1 mM EDTA, 0.5 mM MgCl₂, 1 M sorbitol) was added to stop spheroplasting. Spheroplasted cells (15 μl) were fixed with 30 μl of fixative solution (4% paraformaldehyde, 3.4% sucrose) and lysed with 60 μl 1% lipsol. After addition of 60 μl fixative solution, cells were spread using a glass rod. After drying, the slides were blocked in blocking buffer (0.2% gelatine, 0.5% BSA in PBS) and stained with the following antibodies: anti-HA (3F10, 1:500 dilution, Roche), and anti-Nop1 (1:500 dilution, Encor Biotechnology).

CHEF gel electrophoresis and Southern blotting. Chromosome fragments for CHEF analysis were prepared by restriction digest in agarose plugs. Briefly, 20 ml of meiotic cells were killed by addition of sodium azide (0.1% final w/v), pelleted and stored on ice for the duration of the time course. Cell pellets were washed twice in CHEF-TE (10 mM Tris-HCl (pH 7.5), 50 mM EDTA) and resuspended in 300 μl CHEF-TE. Tubes were individually treated as follows: 4 μl zymolyase T100 (10 mg ml^{–1}) was added and the mix was incubated at 42 °C for 30 s before addition of 500 μl low-melting-point agarose (1.5% SeaPlaque GTG, 125 mM EDTA) at 42 °C. Gel plugs (90 μl) were allowed to harden on ice in disposable plug molds (Bio-Rad) and incubated overnight at 37 °C in 300 μl LET (10 mM Tris (pH 7.5), 500 mM EDTA) per plug. Plugs were deproteinized overnight at 50 °C in 200 μl NDS-PK (LET, 1% N-lauroylsarcosine, 1 mg ml^{–1} proteinase K (Amresco)) per plug. Proteinase K was inactivated by incubating plugs for 1 h at 4 °C in CHEF-TE containing 1 mM PMSF, and plugs were washed three times in CHEF-TE, then digested with XhoI in digestion buffer containing 5 mM spermidine. To analyse the entire rDNA array, digested chromosomes were separated by CHEF gel electrophoresis in 1% agarose in 0.5× TBE, 6 V cm^{–1}, using 60-s pulses for 15 h and 90-s pulses for 9 h. For fine mapping of rDNA insertions and to analyse changes in repeat number, digested chromosome fragments were separated using a 5–20 s ramp over 20 h. For conventional electrophoresis, DNA fragments were separated on 0.6% agarose in 1× TBE and transferred onto Hybond-XL membranes (GE Healthcare) by alkaline transfer. Southern blotting was performed as previously described³² and quantified with a Fujifilm BAS-2500 image reader V1.8 and Multi Gauge V2.2 software.

Probes for Southern analysis. Probe templates for non-rDNA sequences were generated by nested PCR and gel purification. The following probes (SGD coordinates) were used:

YLR164W: XII, 493,432–493,932.

YCR047C: III, 209,361–210,030.

ARS1216: XII, 450,407–451,150.

YLR152C: XII, 443,849–444,910.

NTS1: XhoI and XbaI digest of pRS306-NTS1/2. This probe detects all *NTS1* and *NTS2* sequences (pan-rDNA probe).

rDNA insertion: BciVI digest of pRS306-NTS1/2. This probe specifically detects the plasmid backbone of the *URA3* insertion cassette.

TOM1/YDR457W: IV: 1,370,714–1,371,733

Flow cytometry. At the indicated time points, 150 μ l of meiotic cells were fixed for 2 h at 4 °C after addition of 350 μ l absolute ethanol. Cells were resuspended in 500 μ l of 50 mM sodium citrate containing 0.7 μ l RNase A (30 mg ml⁻¹, Sigma). Cells were incubated for 2 h at 50 °C. 10 μ l proteinase K (Amresco) was added and cells were deproteinated for 2 h at 50 °C. 500 μ l of 50 mM sodium citrate containing 0.2 μ l Sytox Green (Amersham) was added to the cells. Cells were briefly sonicated and analysed using a FACScalibur (Becton Dickinson) flow cytometer. DNA profiles were generated using CellQuest software.

Recombination mapping. To determine crossover recombination rates, cells were sporulated for 24 h in 3 ml SPO and treated with zymolyase (1 mg ml⁻¹ in 1 M sorbitol) to remove ascus walls. Tetrads were dissected by micromanipulation and marker segregation was determined by replica plating on appropriate selective media. For mapping within the rDNA, only tetratypes were used to calculate recombination rates, to avoid distortions originating from non-parental ditypes that were probably the result of previous mitotic recombination.

24. Bell, S. P. *et al.* The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83**, 563–568 (1995).
25. Kobayashi, T. Strategies to maintain the stability of the ribosomal RNA gene repeats—collaboration of recombination, cohesion, and condensation. *Genes Genet. Syst.* **81**, 155–161 (2006).
26. Chernoff, Y. O., Vincent, A. & Liebman, S. W. Mutations in eukaryotic 18S ribosomal RNA affect translational fidelity and resistance to aminoglycoside antibiotics. *EMBO J.* **13**, 906–913 (1994).
27. Peoples, T. L. *et al.* Close, stable homolog juxtaposition during meiosis in budding yeast is dependent on meiotic recombination, occurs independently of synapsis, and is distinct from DSB-independent pairing contacts. *Genes Dev.* **16**, 1682–1695 (2002).
28. James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425–1436 (1996).
29. Blitzblau, H. G. *et al.* Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr. Biol.* **17**, 2003 (2007).
30. Blitzblau, H. G. & Hochwagen, A. Genome-wide detection of meiotic DNA double-strand break hotspots using single-stranded DNA. *Methods Mol. Biol.* **745**, 47–63 (2011).
31. Loidl, J., Nairz, K. & Klein, F. Meiotic chromosome synapsis in a haploid yeast. *Chromosoma* **100**, 221–228 (1991).
32. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).

Engineered reversal of the β -oxidation cycle for the synthesis of fuels and chemicals

Clementina Dellomonaco¹, James M. Clomburg¹, Elliot N. Miller¹ & Ramon Gonzalez^{1,2}

Advanced (long-chain) fuels and chemicals are generated from short-chain metabolic intermediates through pathways that require carbon-chain elongation. The condensation reactions mediating this carbon-carbon bond formation can be catalysed by enzymes from the thiolase superfamily, including β -ketoacyl-acyl-carrier protein (ACP) synthases, polyketide synthases, 3-hydroxy-3-methylglutaryl-CoA synthases, and biosynthetic thiolases¹. Pathways involving these enzymes have been exploited for fuel and chemical production, with fatty-acid biosynthesis (β -ketoacyl-ACP synthases) attracting the most attention in recent years^{2–4}. Degradative thiolases, which are part of the thiolase superfamily and naturally function in the β -oxidation of fatty acids^{5,6}, can also operate in the synthetic direction and thus enable carbon-chain elongation. Here we demonstrate that a functional reversal of the β -oxidation cycle can be used as a metabolic platform for the synthesis of alcohols and carboxylic acids with various chain lengths and functionalities. This pathway operates with

coenzyme A (CoA) thioester intermediates and directly uses acetyl-CoA for acyl-chain elongation (rather than first requiring ATP-dependent activation to malonyl-CoA), characteristics that enable product synthesis at maximum carbon and energy efficiency. The reversal of the β -oxidation cycle was engineered in *Escherichia coli* and used in combination with endogenous dehydrogenases and thioesterases to synthesize *n*-alcohols, fatty acids and 3-hydroxy-, 3-keto- and trans- Δ^2 -carboxylic acids. The superior nature of the engineered pathway was demonstrated by producing higher-chain linear *n*-alcohols ($C \geq 4$) and extracellular long-chain fatty acids ($C > 10$) at higher efficiency than previously reported^{2,4,7–9}. The ubiquitous nature of β -oxidation, aldehyde/alcohol dehydrogenase and thioesterase enzymes has the potential to enable the efficient synthesis of these products in other industrial organisms.

Engineering the β -oxidation cycle as a metabolic platform for the synthesis of advanced fuels and chemicals (Fig. 1) requires the reverse

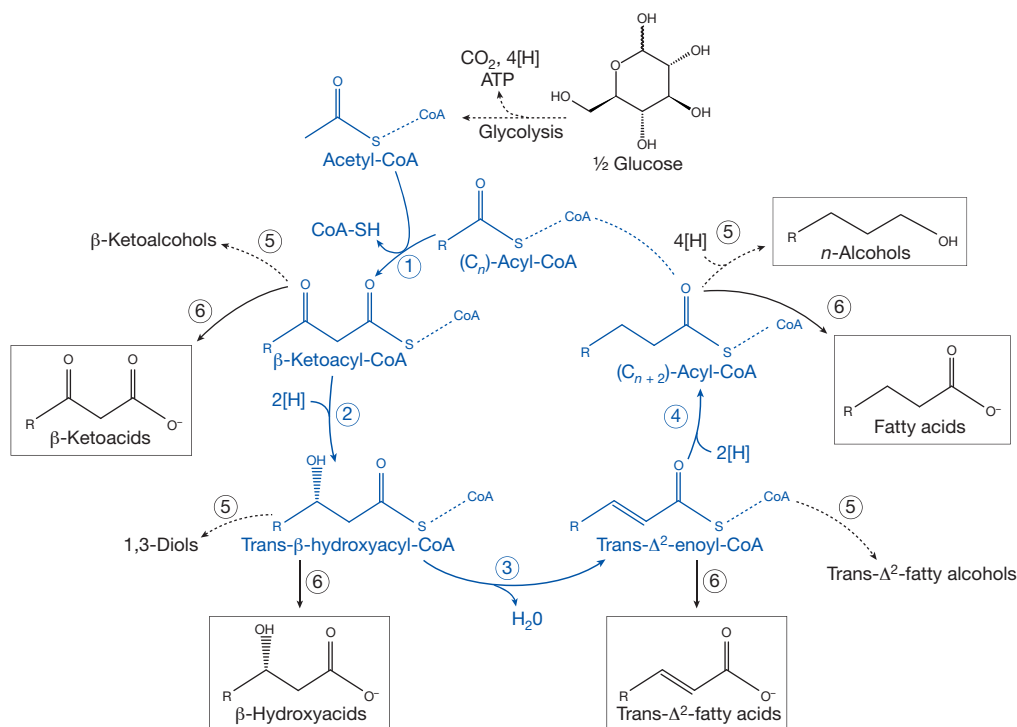


Figure 1 | Functional reversal of the β -oxidation cycle. The following enzymes (gene names in parentheses) are involved in the functional reversal of the β -oxidation cycle engineered in this work (shown in blue): (1) thiolase (*yqeF*, *fadA*); (2) hydroxyacyl-CoA dehydrogenase (*fadB*); (3) enoyl-CoA hydratase (*fadB*); (4) enoyl-CoA reductase (*ydiO*). Each turn of the cycle generates an acyl-CoA that is two carbons longer than the initial acyl-CoA thioester (indicated as $C_n + 2$). Intermediates in the engineered pathway can be converted to a functionally diverse set of molecules using aldehyde-forming acyl-CoA reductases and alcohol dehydrogenases (5) and acyl-CoA

thioesterases (6). Products for which synthesis was demonstrated in this study are boxed. R, side chain attached to the acyl-CoA group of the starter molecule (for example, R = H for acetyl-CoA and R = CH₃ for propionyl-CoA); 2[H], NADH/NADPH/ferredoxin. Dotted lines indicate multiple steps. Dashed lines without arrowheads connect identical metabolites of different chain length. A comparison of *n*-alcohol production through the engineered β -oxidation cycle and the recently proposed fatty-acid biosynthesis pathway² is shown in Supplementary Fig. 1.

¹Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, USA. ²Department of Bioengineering, Rice University, Houston, Texas 77005, USA.

operation of this pathway in the absence of its natural substrate (that is, fatty acids) and presence of a non-fatty-acid carbon source (for example, glucose). Constitutive expression of the *fad* and *ato* regulons, which encode the β -oxidation system in *Escherichia coli*⁵, in the absence of fatty acids was achieved by introducing previously reported *fadR* and *atoC(c)* mutations¹⁰ (Supplementary Table 1). Several operons encoding β -oxidation cycle enzymes are also activated by the cyclic AMP (cAMP) receptor protein (CRP)–cAMP complex and therefore subjected to carbon catabolite repression in the presence of glucose¹¹. This regulatory mechanism was circumvented by replacing the native *crp* gene with a cAMP-independent mutant (*crp**) previously reported to confer a catabolite-derepressed phenotype¹² (Supplementary Table 1). Finally, the *arcA* gene was deleted to prevent ArcA-mediated repression of most operons encoding the β -oxidation cycle¹³.

The genetic modifications outlined above were introduced into wild-type *E. coli* MG1655 (resulting strain named RB01: MG1655 *fadR atoC(c) crp* ΔarcA*) and in a strain in which other fermentative pathways were blocked (resulting strain named RB02: MG1655 *fadR atoC(c) crp* ΔarcA ΔadhE Δpta ΔfrdA*). However, none of the products that can be derived from the β -oxidation cycle intermediates (Fig. 1) was synthesized by these strains (Supplementary Table 2). A limitation on the expression of β -oxidation enzymes was ruled out after confirming the presence of these activities in strain RB02 (Table 1a).

The expression of appropriate termination enzymes is vital to achieve the conversion of the CoA-thioester intermediates in the reversal of the β -oxidation cycle to the desired product(s) (Fig. 1). We first targeted the synthesis of *n*-butanol, as this product can be generated in a one-turn reversal of the β -oxidation cycle. Unlike the significant activity observed for key β -oxidation enzymes, the level of *n*-butanol dehydrogenase in strain RB02 was very low (Table 1a). To address this issue, two endogenous aldehyde/alcohol dehydrogenases with high sequence and structural similarity to clostridial butyraldehyde/butanol dehydrogenase were expressed in strain RB02: that is, L-1,2-propanediol oxidoreductase (FucO) and an aldehyde/alcohol dehydrogenase (YqhD) (Supplementary Table 3). These modifications enabled a functional one-turn reversal of the β -oxidation cycle, as evidenced by the synthesis of *n*-butanol in strains RB02 [*yqhD*+] and RB02 [*fucO*+] (Fig. 2a). Despite the potential for YqhD to catalyse the conversion of butyraldehyde to *n*-butanol¹⁴, overexpression of FucO led to higher concentration of *n*-butanol and lower concentration of the major fermentation by-product, ethanol (Fig. 2a).

The initiation of the reversal of the β -oxidation cycle (that is, conversion of acetyl-CoA to acetoacetyl-CoA) is another key step in the engineered pathway. Therefore, we evaluated the effect of overexpressing *atoB*-encoded acetyl-CoA acetyltransferase, which has high specificity for short-chain acyl-CoA molecules^{15,16} and a predicted acyltransferase (YqeF) with high sequence similarity to AtoB. These modifications also supported a functional reversal of the β -oxidation cycle, as evidenced by the synthesis of *n*-butanol in strains RB02 [*atoB*+] and RB02 [*yqeF*+] (Fig. 2a). The simultaneous overexpression of YqeF and FucO in strain RB02 yielded a high *n*-butanol titre (1.9 g l^{-1}), *n*-butanol yield ($0.19 \text{ g n-butanol per g total glucose consumed}$) and *n*-butanol-to-ethanol ratio ($>5:1$) (Fig. 2a). No *n*-butanol was synthesized upon overexpression of FucO and YqeF in wild-type MG1655 (Supplementary Table 2), demonstrating the requirement of an active reversal of the β -oxidation cycle.

Further reduction in the synthesis of by-product ethanol was realized through deletion of *eutE* and *yqhD*: EutE is an aldehyde dehydrogenase with high sequence similarity to AdhE and YqhD overexpression favoured the synthesis of ethanol over *n*-butanol (Fig. 2a). The resulting strain (RB02 $\Delta yqhD \Delta eutE$ [*yqeF*+*fucO*]) synthesized 2.2 g l^{-1} of *n*-butanol in 24 h at a yield of $0.28 \text{ g n-butanol per g total glucose consumed}$ (Fig. 2a). When grown in a bioreactor using a higher concentration of glucose, this strain produced *n*-butanol at high titre

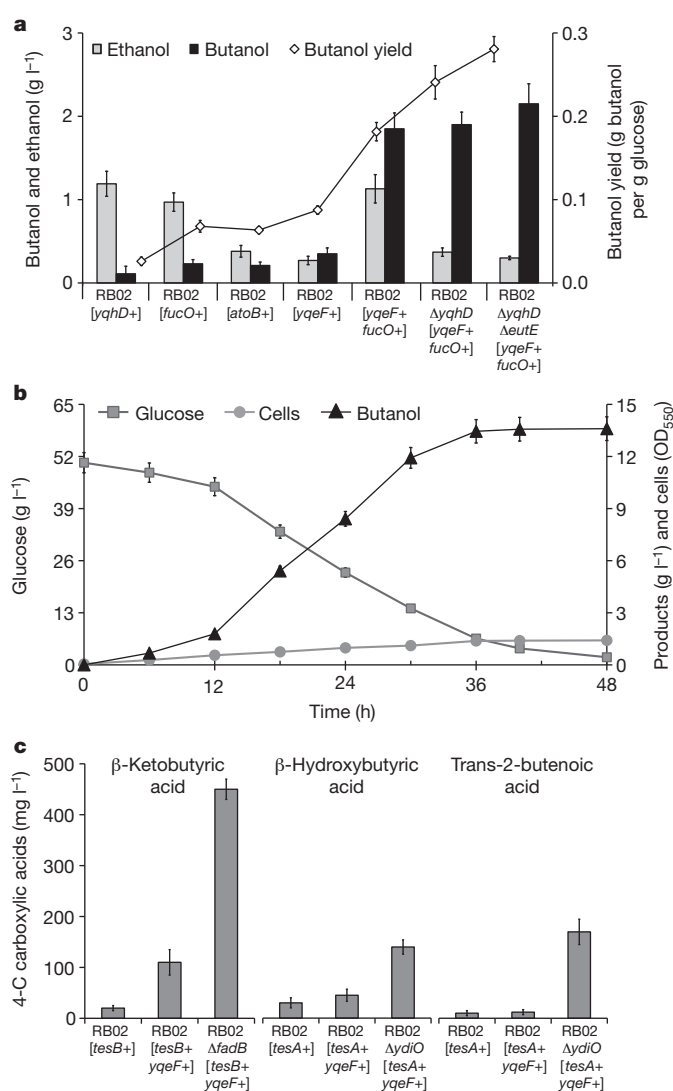


Figure 2 | Engineered one-turn reversal of the β -oxidation cycle for the synthesis of *n*-butanol and short-chain carboxylic acids. **a**, Effect of gene overexpression and knockout (indicated below x axis) on the synthesis of *n*-butanol and ethanol in strain RB02 (*fadR atoC(c) crp* ΔarcA ΔadhE Δpta ΔfrdA*). Experiments were performed at 30 °C for 24 h in shake flasks using glucose (2% w/v) minimal medium. The *n*-butanol yield was calculated as g *n*-butanol per g total glucose consumed. **b**, Kinetics of *n*-butanol production by strain RB02 $\Delta yqhD \Delta eutE$ [*yqeF*+*fucO*]. Cells were cultivated in fermentors containing minimal medium supplemented with 5% (w/v) glucose. The dissolved oxygen was controlled at 5% of saturation, temperature at 30 °C, and pH at 7. **c**, Synthesis of β -ketobutyric acid (left), β -hydroxybutyric acid (centre) and trans-2-butenic acid (right) upon overexpression of thioesterases I (*TesA*) and II (*TesB*) in strains RB02, RB02 $\Delta fadB$ and RB02 $\Delta ydiO$. Experiments were run at 37 °C for 48 h in shake flasks using glucose (1% w/v) minimal medium. Error bars represent the standard deviation of three replicates.

($\sim 14 \text{ g l}^{-1}$), yield ($0.33 \text{ g n-butanol per g total glucose consumed}$) and rate ($\sim 2 \text{ g n-butanol per g cell dry weight per h}$) (Fig. 2b). This performance, which was achieved in the absence of rich nutrients and without importing foreign genes, is better than reported for other organisms engineered for *n*-butanol production (Supplementary Table 4) and also surpasses the *n*-butanol yield and specific productivity reported for native producers¹⁷. The engineered reversal of the β -oxidation cycle operated at a maximum flux of $73.4 \text{ mmol acetyl-CoA per g cell dry weight per h}$ (12–18 h in Fig. 2b), which exceeds the flux reported in the literature for native or engineered fermentative pathways^{18,19}.

A characterization of the engineered β -oxidation pathway was conducted to confirm the identity of the enzymes catalysing each

individual step and their role in *n*-butanol synthesis (Table 1). Enzyme activity measurements showed a high level of expression of key pathway enzymes in strain RB02 [*yqeF*+*fucO*+] but negligible activity in wild-type MG1655 (Table 1a). Gene knockout and gene complementation experiments along with quantification of fermentation products (Table 1b) demonstrated that the primary enzymes involved in the synthesis of *n*-butanol through the engineered reversal of the β -oxidation pathway are (activities shown in parentheses along with reaction numbers, as per Fig. 1): YqeF (predicted acyltransferase: reaction (1)); FadB (3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase: reactions (2) and (3), respectively); YdiO (predicted acyl-CoA dehydrogenase/enoyl-CoA reductase: reaction (4)), and MhpF and FucO (aldehyde-forming acyl-CoA reductase and *n*-butanol dehydrogenase, respectively: reaction (5)).

YdiO is proposed to catalyse the reduction of enoyl-CoA to acyl-CoA (reaction (4)). The reverse of this reaction is catalysed by FadE and is the only irreversible step in the catabolic operation of the β -oxidation cycle⁵. In agreement with our proposal, deletion of *ydiO* in strain RB02 [*yqeF*+*fucO*+] completely abolished *n*-butanol synthesis (Table 1b). Although *ydiO* was previously proposed to encode an acyl-CoA dehydrogenase that would replace FadE during the anaerobic catabolism of fatty acids²⁰, a sequence comparison between YdiO and *E. coli* proteins does not reveal a significant similarity to FadE (Supplementary Table 5). In contrast, YdiO shares high homology with crotonobetainyl-CoA reductase (CaiA; Supplementary Table 5). CaiA catalyses the reduction of crotonobetainyl-CoA to γ -butyrobetainyl-CoA²¹, a reaction similar to that catalysed by YdiO in the reversal of the β -oxidation cycle. Moreover, the operon *fixABCX* is required for the transfer of electrons to CaiA^{22,23} and encodes flavoproteins and ferredoxin with high sequence similarity to YdiQRST (Supplementary Table 5). This analysis suggests that ferredoxin and flavoproteins encoded by *ydiQRST*²⁰ are involved in the transfer of electrons to YdiO during the reduction of enoyl-CoA to acyl-CoA. Standard Gibbs energy calculations reveal that the engineered reversal of the β -oxidation cycle is thermodynamically feasible if ferredoxin is the source of reducing power for the conversion of enoyl-CoA to acyl-CoA

(Supplementary Table 6). We then propose that the reduction of enoyl-CoA to acyl-CoA is mediated by YdiO-YdiQRST.

The engineered reversal of the β -oxidation cycle generates a diverse set of CoA thioester intermediates that can be converted to the corresponding alcohols and carboxylic acids (Fig. 1). To illustrate product synthesis from intermediates other than acyl-CoA, we used thioesterase I (TesA)²⁴ and thioesterase II (TesB)²⁵ as termination enzymes. Small amounts of 3-hydroxybutyric, 3-ketobutyric and trans-2-butenic acids were produced when these thioesterases were overexpressed in strain RB02 (Fig. 2c). The level of these products was significantly increased by simultaneous overexpression of thioesterase and *yqeF*-encoded short-chain acyltransferase (Fig. 2c). Further increases in titre were realized upon deletion of *fadB* ($\sim 500 \text{ mg l}^{-1}$ 3-ketobutyric acid) and *ydiO* ($\sim 150 \text{ mg l}^{-1}$ and 200 mg l^{-1} of 3-hydroxybutyric and trans-2-butenic acids, respectively) (Fig. 2c).

The operation of multiple cycles of the engineered reversal of the β -oxidation pathway, and hence the synthesis of CoA-thioester intermediates (and products) of longer chain length ($C > 4$), can be facilitated by the overexpression of FadA, a 3-ketoacyl-CoA thiolase that is part of the β -oxidation complex (FadBA) and which possesses broad chain length specificity²⁶. Overexpression of FadBA in conjunction with thioesterases (TesA, TesB, FadM²⁷ or YciA²⁸) in strain RB03 (RB02 $\Delta yqhD \Delta fucO \Delta fadD$) resulted in the accumulation of long-chain fatty acids in the extracellular medium (Fig. 3a). The *fadD* knockout in strain RB03 prevents re-utilization of the synthesized fatty acids. The choice of thioesterase allowed control over both length and functionality of the fatty-acid side chain. For example, C16 and C18 saturated fatty acids were the only products when FadM was overexpressed, whereas YciA and TesA overexpression supported the synthesis of 3-hydroxy (C14:3OH) and unsaturated (C18:1) fatty acids, respectively (Fig. 3a). When grown in a bioreactor using a higher initial concentration of glucose, strain RB03 [*fadBA*-*fadM*+] produced long-chain extracellular fatty acids at high titre ($\sim 7 \text{ g l}^{-1}$) and yield (0.28 g fatty acids per g total glucose consumed) using a mineral salts medium without rich nutrients (Fig. 3b). These results are better than reported previously using an engineered fatty-acid biosynthesis pathway²

Table 1 | Characterization of the engineered reversal of the β -oxidation cycle during *n*-butanol synthesis

Strain	Enzyme activity ($\mu\text{mol per mg protein per min}$) \pm s.d.			
	THL	HBD	CRT	BDH
(a) Activities of β-oxidation and butanol dehydrogenase enzymes in wild-type and engineered strains				
MG1655	ND	0.002 \pm 0.000	ND	0.014 \pm 0.001
RB02 (<i>fadR</i> <i>atoC</i> (c) <i>crp</i> * Δ <i>arcA</i> Δ <i>pta</i> Δ <i>adhE</i> Δ <i>frdA</i>)	0.310 \pm 0.079	0.304 \pm 0.032	0.339 \pm 0.049	0.004 \pm 0.002
RB02 [<i>yqeF</i> + <i>fucO</i> +]	0.498 \pm 0.036	0.292 \pm 0.013	0.334 \pm 0.017	0.298 \pm 0.020
Strain	Butanol produced		Glucose used	Cell growth
	Yield (g g^{-1})	Concentration (g l^{-1})	(g l^{-1})	(g l^{-1})
(b) Butanol synthesis, glucose utilization and cell growth in strain RB02 and its derivatives*				
RB02 [<i>yqeF</i> + <i>fucO</i> +]	0.182	1.85	10.19	0.72
Reaction (1): YqeF (predicted acyltransferase)				
RB02 $\Delta yqeF$ [<i>fucO</i> +]	0.019	0.10	5.19	0.32
RB02 [<i>fucO</i> +]	0.063	0.23	3.66	0.41
RB02 $\Delta yqeF$ [<i>fucO</i> + <i>yqeF</i> +]	0.159	1.11	7.00	0.43
Reactions (2) and (3): FadB (3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase)				
RB02 $\Delta fadB$ [<i>yqeF</i> + <i>fucO</i> +]	ND	ND	2.96	0.19
RB02 $\Delta fadB$ [<i>yqeF</i> + <i>fadB</i> +]	0.157	0.92	5.86	0.52
RB02 $\Delta fadB$ [<i>yqeF</i> +]	ND	ND	1.84	0.11
Reaction (4): YdiO (predicted acyl-CoA dehydrogenase/enoyl-CoA reductase)				
RB02 $\Delta ydiO$ [<i>yqeF</i> + <i>fucO</i> +]	0.038	0.12	3.14	0.18
RB02 $\Delta ydiO$ [<i>yqeF</i> + <i>ydiO</i> +]	0.163	0.97	5.94	0.57
RB02 $\Delta ydiO$ [<i>yqeF</i> +]	0.018	0.10	1.66	0.13
Reaction (5): MhpF (aldehyde-forming acyl-CoA reductase) and FucO (<i>n</i> -butanol dehydrogenase)				
RB02 $\Delta fucO$ [<i>yqeF</i> +]	0.015	0.04	2.75	0.38
RB02 $\Delta fucO$ [<i>yqeF</i> + <i>fucO</i> +]	0.151	0.85	5.64	0.51
RB02 [<i>yqeF</i> +]	0.088	0.35	4.00	0.68
RB02 $\Delta mhpF$ [<i>yqeF</i> + <i>fucO</i> +]	ND	ND	0.66	0.07

BDH, butanol dehydrogenase; CRT, crotonase; HBD, hydroxy-acyl-CoA dehydrogenase; THL, thiolase.

ND, not detectable (minimum detectable activity was 0.001 $\mu\text{mol per mg protein per min}$; minimum detection level for butanol was 5.84 mg l^{-1}).

*Experiments were run for 24 h in shake flasks using glucose (2% w/v) minimal medium at 30 °C.

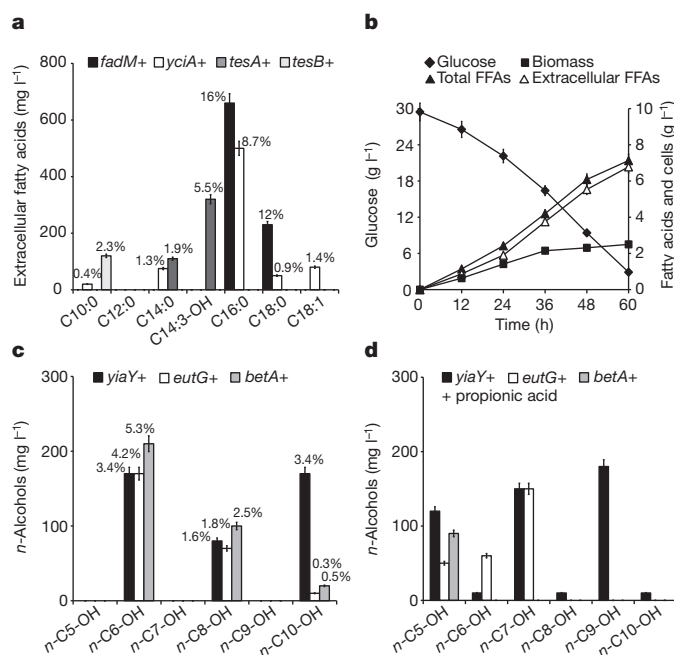


Figure 3 | Synthesis of higher-chain ($C > 4$) carboxylic acids and n -alcohols through the engineered reversal of the β -oxidation cycle. **a**, Accumulation of long-chain ($C > 10$) free fatty acids in the extracellular medium of strain RB03 [*fadBA*+] upon overexpression of different thioesterases (FadM, YciA, TesA, TesB). Product yield is shown above the bars (g free fatty acid per g total glucose consumed $\times 100$). Experiments were run at 37 °C for 48 h in shake flasks using glucose (2% w/v) minimal medium. **b**, Kinetics of fatty-acid synthesis by strain RB03 [*fadBA-fadM*]. Cells were cultivated in fermentors using glucose (3% w/v) minimal medium. The dissolved oxygen was controlled at 2% of saturation, 37 °C, pH 7. FFAs, free fatty acids. **c**, Synthesis of n -alcohols in strain RB03 [*fadBA*+] upon overexpression of alcohol dehydrogenases (YiaY, BetA and EutG). Product yield is shown above the bars (g n -alcohol per g total glucose consumed $\times 100$). Experiments were run at 37 °C for 48 h in shake flasks using glucose (2% w/v) minimal medium. **d**, Effect of alcohol dehydrogenase overexpression (YiaY, BetA and EutG) on the chain-length distribution of n -alcohols synthesized by strain RB03 [*fadBA*+] in the presence of 0.5 g l⁻¹ propionate. Experiments were conducted as described in c. Error bars represent the standard deviation of three replicates.

(Supplementary Table 4). No production of extracellular free fatty acids was observed upon overexpression of FadM in strain MG1655 Δ *adhE* Δ *pta* Δ *frdA* Δ *fadD* (Supplementary Table 2C), demonstrating the requirement of an active reversal of the β -oxidation cycle. Measurements of total free fatty acids (that is, extracellular plus intracellular) in strain RB03 [*fadBA-fadM*+] and the corresponding controls showed that the engineered reversal of the β -oxidation cycle contributed to the synthesis of 90–95% of the total free fatty acids (Supplementary Table 2C).

The synthesis of longer-chain ($C > 4$) n -alcohols was also demonstrated by overexpressing the appropriate termination enzymes (Fig. 3c). We identified native enzymes that could serve as potential surrogates for the aldehyde-forming acyl-CoA reductases and alcohol dehydrogenases present in organisms that synthesize higher-chain linear n -alcohols (Supplementary Table 3). The product titre (0.33 g l⁻¹) and yield (8.3% w/w) achieved upon overexpression of YiaY were higher than previously reported^{2,29} (Supplementary Table 4). Synthesis of odd-chain n -alcohols was demonstrated by supplementing the medium with propionate as the precursor of propionyl-CoA ($R = \text{CH}_3$ in Fig. 1). A clear shift in the distribution of n -alcohols was observed: odd-chain alcohols 1-pentanol, 1-heptanol and 1-nonanol appeared as fermentation products and the synthesis of even-chain alcohols significantly decreased (Fig. 3d).

The functional reversal of the β -oxidation cycle engineered in this work represents a novel and efficient platform for the synthesis of

advanced fuels and chemicals. Its superior nature is illustrated in the following balanced equation for the synthesis of n -alcohols from glucose: $n/4 \text{ C}_6\text{H}_{12}\text{O}_6 \rightarrow \text{C}_n\text{H}_{n+2}\text{O} + n/2 \text{ CO}_2 + (n/2 - 1) \text{ H}_2\text{O} + n/2 \text{ ATP}$, with n being the chain length of the n -alcohol (Fig. 1). As can be seen in this balance, the engineered pathway has the potential to achieve the maximum theoretical yield of n -alcohols on glucose (66.7%, C-mole basis) and generates 1 ATP per each two-carbon unit incorporated into the n -alcohol molecule. This ATP yield is equivalent to that of efficient homo-fermentative pathways found in nature such as ethanol³⁰ and lactic acid¹⁹ fermentations. The high efficiency of the engineered reversal of the β -oxidation cycle is possible because it directly uses acetyl-CoA as the donor of two-carbon units during chain elongation (as opposed to first requiring ATP-dependent activation to malonyl-CoA) and it functions with acyl-CoA intermediates, which are the precursors of n -alcohols and other important products (Fig. 1).

The synthesis of n -alcohols through alternative metabolic routes, such as the fatty-acid biosynthesis² and keto-acid²⁹ pathways, is less efficient. The use of the fatty acid biosynthesis pathway results in the net consumption of 1 ATP per molecule of n -alcohol synthesized (Supplementary Fig. 1). This inefficiency is due to the consumption of ATP in the synthesis of malonyl-ACP, the donor of two-carbon units for chain elongation, and the use of acyl-ACP intermediates, which need to be converted to free acids and acylated (ATP-consuming step) before their reduction to alcohols can be achieved (Supplementary Fig. 1). The keto-acid pathway²⁹ is also less efficient than the reversal of the β -oxidation cycle: for example, the maximum theoretical yield of n -hexanol, the highest-chain linear n -alcohol reported with the keto-acid pathway, is only 50% C-mole (2 glucose \rightarrow n -hexanol + ATP + 2[H] + 6 CO₂).

METHODS SUMMARY

Strains and plasmids. Wild-type K12 *Escherichia coli* strain MG1655 was used as the host to implement metabolic engineering strategies. Genetic manipulations, including construction of plasmids, are described in the Methods. All resulting strains and plasmids are listed in Supplementary Table 7.

Culture medium, cultivation conditions and analytical methods. The mineral salts medium described in Methods was used. Ampicillin (100 $\mu\text{g ml}^{-1}$), kanamycin (50 $\mu\text{g ml}^{-1}$), chloramphenicol (34 $\mu\text{g ml}^{-1}$), anhydrotetracycline (100 ng ml⁻¹) and isopropyl β -D-1-thiogalactopyranoside (IPTG; 0.1 mM) were included when appropriate. Several colonies were used to inoculate 30 ml of medium in 50 ml baffled flasks and incubated at 37 °C in a rotary shaker (200 r.p.m.) for 24 h. An appropriate amount of pellet from this pre-culture was used to inoculate 30 ml of fresh medium in 50 ml baffled flasks to a target initial optical density at 550 nm of 0.05. These cultures were then incubated at 30 °C in a rotary shaker (200 r.p.m.) for 24 h, unless otherwise stated. The same inoculation procedure was used for additional fermentations conducted in a SixFors multi-fermentation system (Infors HT) with an air flowrate of 1.5 Nl h⁻¹ and control of dissolved oxygen, pH, and temperature at specified values. Metabolite identification and quantification by nuclear magnetic resonance (NMR), gas chromatography-mass spectrometry (GC-MS), gas chromatography-flame ionization detection (GC-FID), and high-performance liquid chromatography (HPLC) are described in Methods. Product yields were calculated as grams of product synthesized per total grams of glucose consumed. All values reported as different are significant with a P -value lower than 0.05.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 January; accepted 23 June 2011.

Published online 10 August 2011.

- Haapalainen, A. M., Merilainen, G. & Wierenga, R. K. The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem. Sci.* **31**, 64–71 (2006).
- Steen, E. J. et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559–562 (2010).
- Handke, P., Lynch, S. A. & Gill, R. T. Application and engineering of fatty acid biosynthesis in *Escherichia coli* for advanced fuels and chemicals. *Metab. Eng.* **13**, 28–37 (2010).
- Lennen, R. M., Braden, D. J., West, R. M., Dumesic, J. A. & Pfleger, B. F. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnol. Bioeng.* **106**, 193–202 (2010).

5. Clark, D. P. & Cronan, J. E. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhart, F. C.) Ch. 3.4.4 343–357 (American Society for Microbiology, 2005).
6. Poirier, Y., Antonenkov, V. D., Glumoff, T. & Hiltunen, J. K. Peroxisomal β -oxidation—A metabolic pathway with multiple functions. *Biochim. Biophys. Acta* **1763**, 1413–1426 (2006).
7. Shen, C. R. & Liao, J. C. Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* **10**, 312–320 (2008).
8. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
9. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nature Chem. Biol.* **7**, 222–227 (2011).
10. Dellomonaco, C., Rivera, C., Campbell, P. & Gonzalez, R. Engineered respiro-fermentative metabolism for the production of biofuels and biochemicals from fatty acid-rich feedstocks. *Appl. Environ. Microbiol.* **76**, 5067–5078 (2010).
11. Deutscher, J. The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* **11**, 87–93 (2008).
12. Eppler, T. & Boos, W. Glycerol-3-phosphate-mediated repression of *malT* in *Escherichia coli* does not require metabolism, depends on enzyme IIA(Glc) and is mediated by cAMP levels. *Mol. Microbiol.* **33**, 1221–1231 (1999).
13. Cho, B. K., Knight, E. M. & Palsson, B. O. Transcriptional regulation of the *fad* regulon genes of *Escherichia coli* by ArcA. *Microbiology-Sgm* **152**, 2207–2219 (2006).
14. Jarboe, L. Y. qhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol.* **89**, 249–257 (2010).
15. Feigenbaum, J. & Schulz, H. Thiolases of *Escherichia coli*: Purification and chain length specificities. *J. Bacteriol.* **122**, 407–411 (1975).
16. Jenkins, L. S. & Nunn, W. D. Genetic and molecular characterization of the genes involved in short-chain fatty acid degradation in *Escherichia coli*: the *ato* system. *J. Bacteriol.* **169**, 42–52 (1987).
17. Lee, S. Y. *et al.* Fermentative butanol production by clostridia. *Biotechnol. Bioeng.* **101**, 209–228 (2008).
18. Gonzalez, R., Tao, H., Shanmugam, K. T., York, S. W. & Ingram, L. O. Global gene expression differences associated with changes in glycolytic flux and growth rate in *Escherichia coli* during the fermentation of glucose and xylose. *Biotechnol. Prog.* **18**, 6–20 (2002).
19. Papagianni, M., Avramidis, N. & Filiouis, G. Glycolysis and the regulation of glucose transport in *Lactococcus lactis* spp. *lactis* in batch and fed-batch culture. *Microb. Cell Fact.* **6**, doi:10.1186/1475-2859-6-16 (2007).
20. Campbell, J. W., Morgan-Kiss, R. M. & Cronan, J. E. A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic β -oxidation pathway. *Mol. Microbiol.* **47**, 793–805 (2003).
21. Eichler, K., Bourgis, F., Buchet, A., Kleber, H. P. & Mandrandberthelot, M. A. Molecular characterization of the *cai* operon necessary for carnitine metabolism in *Escherichia coli*. *Mol. Microbiol.* **13**, 775–786 (1994).
22. Walt, A. & Kahn, M. L. The *fixA* and *fixB* genes are necessary for anaerobic carnitine reduction in *Escherichia coli*. *J. Bacteriol.* **184**, 4044–4047 (2002).
23. Eichler, K., Buchet, A., Bourgis, F., Kleber, H. P. & Mandrandberthelot, M. A. The *fix* *Escherichia coli* region contains four genes related to carnitine metabolism. *J. Basic Microbiol.* **35**, 217–227 (1995).
24. Cho, H. S. & Cronan, J. E. *Escherichia coli* thioesterase I, molecular cloning and sequencing of the structural gene and identification as a periplasmic enzyme. *J. Biol. Chem.* **268**, 9238–9245 (1993).
25. Nie, L., Ren, Y., Janakiraman, A., Smith, S. & Schulz, H. A novel paradigm of fatty acid β -oxidation exemplified by the thioesterase-dependent partial degradation of conjugated linoleic acid that fully supports growth of *Escherichia coli*. *Biochemistry* **47**, 9618–9626 (2008).
26. Yang, S. Y., Yang, X. Y. H., Healylouie, G., Schulz, H. & Elzinga, M. Primary structure of 3-ketoacyl-coenzyme A thiolase from *Escherichia coli* and the structural organization of the *fadAB* operon. *J. Biol. Chem.* **265**, 10424–10429 (1990).
27. Feng, Y. & Cronan, J. E. A new member of the *Escherichia coli* *fad* regulon: Transcriptional regulation of *fadM* (*ybaW*). *J. Bacteriol.* **191**, 6320–6328 (2009).
28. Zhuang, Z. H. *et al.* Divergence of function in the hot dog fold enzyme superfamily: The bacterial thioesterase YciA. *Biochemistry* **47**, 2789–2796 (2008).
29. Zhang, K. C., Sawaya, M. R., Eisenberg, D. S. & Liao, J. C. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl Acad. Sci. USA* **105**, 20653–20658 (2008).
30. Osman, Y. A., Conway, T., Bonetti, S. J. & Ingram, L. O. Glycolytic flux in *Zymomonas mobilis*: enzyme and metabolite levels during batch fermentation. *J. Bacteriol.* **169**, 3726–3736 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Erni and H. Mori for providing research materials, S. Moran and J. F. Fallas Valverde for assistance with NMR techniques, and D. A. Castillo-Rivera, B. Wilson, S. P. T. Matsuda, M. Li and K.-Y. San for assistance with GC-MS techniques. R.G. thanks N. E. Gonzalez, B. C. Gutierrez and M. D. Diaz for their continued support.

Author Contributions R.G. conceived the work. C.D. and R.G. designed the experiments. C.D. conducted all strain characterization experiments. C.D. performed the *in silico* analyses. C.D., J.M.C. and E.N.M. constructed the strains. J.M.C. performed enzyme assays and the thermodynamic analysis. R.G. and C.D. drafted the manuscript. All authors read, edited and approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.G. (Ramon.Gonzalez@rice.edu).

METHODS

Strains and plasmids. Gene knockouts were introduced in MG1655 and its derivatives by P1 phage transduction as previously described³¹. The native cAMP receptor protein (CRP) was replaced with a cAMP-independent CRP (CRP*)³² by introducing the *crp** gene into *crp* deletion mutants via P1 phage transduction using a lysate from strain ET25 (ref. 13), followed by selection on (1% w/v) glycerol minimal media plates with tetracycline. Construction of plasmids (medium-copy pTrcHis2A, Invitrogen, and low-copy pZSblank (ref. 31)) is described in the 'Genetic methods' section. All resulting strains and plasmids are listed in Supplementary Table 7.

Reagents. Chemicals were obtained from Fisher Scientific and Sigma-Aldrich.

Culture medium. The minimal medium designed previously³³ with Na₂HPO₄ in place of K₂HPO₄ and supplemented with 20 g l⁻¹ glucose, 40 g l⁻¹ calcium carbonate, 100 µM FeSO₄, 5 mM calcium pantothenate, 3.96 mM Na₂HPO₄, 5 mM (NH₄)₂SO₄ and 30 mM NH₄Cl was used. Fermentations conducted in the SixFors multi-fermentation system also included 1 mM betaine. Ampicillin (100 µg ml⁻¹), kanamycin (50 µg ml⁻¹), chloramphenicol (34 µg ml⁻¹), anhydrotetracycline (100 ng ml⁻¹) and isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.1 mM) were included when appropriate. Several colonies were used to inoculate 30 ml of medium in 50 ml baffled flasks and incubated at 37 °C in a rotary shaker (200 r.p.m.) for 24 h. An appropriate amount of pellet from this pre-culture was used to inoculate 30 ml of fresh medium in 50 ml baffled flasks to a target initial optical density at 550 nm of 0.05. These cultures were then incubated at 30 °C in a rotary shaker (200 r.p.m.) for 24 h, unless otherwise stated.

Genetic methods. Standard recombinant DNA procedures were used for gene cloning, plasmid isolation and electroporation³⁴. Manufacturer protocols and standard methods were followed for DNA purification (Qiagen), restriction endonuclease digestion (New England Biolabs) and DNA amplification (Stratagene and Invitrogen). For plasmid construction, genes were amplified from MG1655 genomic DNA using primers designed to create 15 bp of homology on each end of the gene insert for subsequent recombination into the desired plasmid. Plasmids were linearized using restriction endonuclease digestion, then recombined with the appropriate gene(s) using an In-Fusion Dry-Down PCR Cloning Kit (Clontech) and subsequently used to transform chemically competent Fusion Blue cells (Clontech). Transformants that grew on LB plates containing the appropriate antibiotic were struck for isolation, and then subjected to preliminary screening by PCR. Colonies passing preliminary inspection were then individually grown for plasmid purification. Purified plasmids were confirmed to have the appropriate insert both by PCR as well as restriction endonuclease digest verification. Plasmids in each case include the plasmid-based promoter, a ribosomal binding site for each gene, MG1655 gene(s), and a plasmid-based terminator. Resulting plasmids are listed in Supplementary Table 7.

Sequencing of regulators. Chromosomal DNA from strain RB02 served as the template for PCR using Phusion High-Fidelity DNA polymerase (Finnzymes). PCR products, containing the gene(s) of interest as well as approximately 500 bp upstream of each gene, were purified using a QIAquick PCR purification kit (Qiagen) and sent to Lone Star Labs for sequencing. Sequencing results were analysed using Sequence Scanner v1.0 (Applied Biosystems) and Serial Cloner v2.1 (Serial Basics, <http://serialbasics.free.fr/Home/Home.html>).

Metabolite identification. The identity of metabolic products was determined through one-dimensional proton nuclear magnetic resonance (NMR) spectroscopy. 60 µl of D₂O and 1 µl of 600 mM NMR internal standard TSP [3-(trimethylsilyl) propionic acid-D₄, sodium salt] were added to 540 µl of the sample (culture supernatant). The resulting solution was then transferred to a 5-mm NMR tube, and one-dimensional proton NMR spectroscopy was performed at 25 °C in a Varian 500-MHz Inova spectrometer equipped with a Penta probe (Varian) using the following parameters: 8,000-Hz sweep width, 2.8-s acquisition time, 256 acquisitions, 6.3 µs pulse width, 1.2-s pulse repetition delay, and pre-saturation for 2 s. The resulting spectrum was analysed using FELIX 2001 software (Accelrys Software). Peaks were identified by their chemical shifts and J-coupling values, which were obtained in separate experiments in which samples were spiked with metabolite standards (2 mM final concentration).

Identification of *n*-alcohols was conducted through GC-MS following a modification of the method reported previously³⁵. The analysis was performed on an Agilent 6890 GC/5973 MS (Agilent Technologies) instrument with a HP-5ms capillary column (30 m × 0.25 mm × 0.25 µm). 1 ml of supernatant of culture broth was extracted with 500 µl of GC standard grade *n*-hexane (Fluka) and 1-heptanol (Sigma-Aldrich) was used as internal standard. 0.5 µl of the extracted sample was injected using a 20:1 split at 250 °C. The oven temperature was initially held at 75 °C for 2 min and then raised with a gradient of 5 °C per min to 280 °C and held for 2 min. Helium (Matheson Tri-Gas) was used as the carrier gas with a 14-lb in⁻² inlet pressure. The injector and detector were maintained at 255 °C.

Identification of fatty acids was performed on a Shimadzu Auto-System GC 2010 (Shimadzu) equipped with a DB-5MS capillary column (30 m × 0.25 mm × 0.25 µm) and coupled with a QP-2010 mass detector. The following method was used: an initial temperature of 50 °C was held for 2 min and then ramped to 220 °C at 4 °C per min and held for 10 min (ref. 2). Extraction and derivatization procedures are described in the 'Metabolite quantification' section.

Metabolite quantification. The quantification of glucose, organic acids, ethanol and butanol was conducted by high-performance liquid chromatography (HPLC). Samples (culture supernatant) were analysed with ion-exclusion HPLC using a Shimadzu Prominence SIL 20 system (Shimadzu Scientific Instruments) equipped with an HPX-87H organic acid column (Bio-Rad) with operating conditions to optimize peak separation (0.3 ml min⁻¹ flowrate, 30 mM H₂SO₄ mobile phase, column temperature 42 °C)³⁶.

Quantification of longer chain (*C* > 4) *n*-alcohols was conducted through gas chromatography (GC) in a Varian CP-3800 gas chromatograph (Varian Associates) equipped with a flame ionization detector (GC-FID). Sample extraction procedure was as described above. The separation of alcohol compounds was carried out using a FactorFour VF-5ht column (15 m, 0.32 µm internal diameter, 0.10 µm film thickness; Varian Associates). The oven temperature was initially held at 40 °C for 1 min and then raised with a gradient of 30 °C min⁻¹ to 130 °C and held for 4 min. The temperature was then raised with a gradient of 15 °C min⁻¹ to 230 °C and held for 4 min. Helium (1 ml min⁻¹, Matheson Tri-Gas) was used as the carrier gas. The injector and detector were maintained at 250 °C. A 0.5-µl sample was injected in splitless injection mode.

For total free fatty acid quantification, 5 ml of culture was placed in a 15-ml Corning CentriStar centrifuge tube (Corning Inc.) in an ice bath and sonicated using a Branson Sonifier 450 (Emerson Industrial Automation) for 5 min at a duty cycle of 30% and a power setting of 1. The tubes were then centrifuged at 12,000g for 5 min and the appropriate volume of supernatant recovered. For extracellular free fatty-acid analysis, the supernatant from non-sonicated culture samples was used instead. 2 ml of supernatant were acidified with 200 µl of acetic acid, supplemented with 100 mg of octanoic acid as an internal standard, and extracted with 2 ml of *n*-hexane-chloroform (4:1, v/v)³⁷. The organic layer was nitrogen evaporated to near dryness, re-dissolved in 1 ml of a mixture of chloroform:methanol:hydrochloric acid (10:1:1, vol/vol/vol)¹⁰ and incubated in a sealed vial at 90 °C for 1 h. Fatty-acid methyl esters were extracted with 1 ml of *n*-hexane-chloroform (4:1, v/v) after addition of 500 µl of 0.9% (wt/vol) NaCl in water. Gas-liquid chromatography was performed with 1 µl of the *n*-hexane:chloroform solution on a Varian FactorFour VF-5ht capillary column (Varian Associates, 15 m length, 0.32 µm diameter and 0.1 µm film thickness) according to the following method: 50 °C held for 1 min, 30 °C min⁻¹ to 160 °C, 15 °C min⁻¹ to 200 °C, 200 °C held for 1.5 min, 10 °C min⁻¹ to 225 °C, and 225 °C held for 15 min (ref. 10).

Enzyme assays. For measurement of enzymatic activities, cells from 24-h shake flask cultures were washed twice with 9 g l⁻¹ sodium chloride under anaerobic conditions and stored at -80 °C until use. Cell extracts for all assays were prepared as follows under anaerobic conditions: 40 units of OD_{550 nm} were re-suspended in 1 ml of 100 mM Tris-HCl buffer (pH 7.0) with 1 mM DTT. After cellular disruption using a Disruptor Genie (Scientific Industries), cellular debris was removed by centrifugation (13,000g, 4 °C, 10 min) and the supernatant used as cell extract. Absorbance changes for all assays were monitored in a Biomate 5 spectrophotometer (Thermo Scientific). The linearity of reactions (protein concentration and time) was established for all assays and the non-enzymatic rates were subtracted from the observed initial reaction rates. Enzymatic activities are reported as µmol of substrate per min per mg of cell protein and represent averages for at least three cell preparations. Protein concentration was measured using the Bradford Assay Reagent (Thermo Scientific) with BSA as a standard.

Acetyl-CoA acetyltransferase (THL) activity was determined using acetoacetyl-CoA and CoA as substrates, and the decrease in acetoacetyl-CoA concentration was measured at 303 nm (ref. 38). β-Hydroxybutyryl-CoA dehydrogenase activity was measured at 340 nm by monitoring the decrease in NADH concentration resulting from β-hydroxybutyryl-CoA formation from acetoacetyl-CoA (ref. 39). Crotonase activity was measured by monitoring the decrease in crotonyl-CoA concentration at 263 nm, which results from β-hydroxybutyryl-CoA formation from crotonyl-CoA (ref. 39). To measure butanol dehydrogenase activity, the decrease in NAD(P)H concentration resulting from butanol formation from butyraldehyde is monitored at 340 nm under anaerobic conditions at 30 °C (ref. 40).

31. Yazdani, S. S. & Gonzalez, R. Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. *Metab. Eng.* **10**, 340–351 (2008).
32. Cirino, P. C., Chin, J. W. & Ingram, L. O. Engineering *Escherichia coli* for xylitol production from glucose-xylose mixtures. *Biotechnol. Bioeng.* **95**, 1167–1176 (2006).
33. Neidhart, F. C., Bloch, P. L. & Smith, D. F. Culture media for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).

34. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, 2001).
35. Atsumi, S. *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
36. Dharmadi, Y. & Gonzalez, R. A better global resolution function and a novel iterative stochastic search method for optimization of high-performance liquid chromatographic separation. *J. Chromatogr. A* **1070**, 89–101 (2005).
37. Lalman, J. A. & Bagley, D. M. Extracting long-chain fatty acids from a fermentation medium. *J. Am. Oil Chem. Soc.* **81**, 105–110 (2004).
38. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Thiolase from *Clostridium acetobutylicum* ATCC824 and its role in the synthesis of acids and solvents. *Appl. Environ. Microbiol.* **54**, 2717–2722 (1988).
39. Hartmanis, M. G. N. & Gatenbeck, S. Intermediary metabolism in *Clostridium acetobutylicum*—levels of enzymes involved in the formation of acetate and butyrate. *Appl. Environ. Microbiol.* **47**, 1277–1283 (1984).
40. Durre, P., Kuhn, A., Gottwald, M. & Gottschalk, G. Enzymatic investigations on butanol dehydrogenase and butyraldehyde dehydrogenase in extracts of *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* **26**, 268–272 (1987).

Modulation of Rab GTPase function by a protein phosphocholine transferase

Shaeri Mukherjee^{1*}, Xiaoyun Liu^{1*}, Kohei Arasaki¹, Justin McDonough¹, Jorge E. Galán¹ & Craig R. Roy¹

The intracellular pathogen *Legionella pneumophila* modulates the activity of host GTPases to direct the transport and assembly of the membrane-bound compartment in which it resides^{1–6}. *In vitro* studies have indicated that the *Legionella* protein DrrA post-translationally modifies the GTPase Rab1 by a process called AMPylation⁷. Here we used mass spectrometry to investigate post-translational modifications to Rab1 that occur during infection of host cells by *Legionella*. Consistent with *in vitro* studies, DrrA-mediated AMPylation of a conserved tyrosine residue in the switch II region of Rab1 was detected during infection. In addition, a modification to an adjacent serine residue in Rab1 was discovered, which was independent of DrrA. The *Legionella* effector protein AnkX was required for this modification. Biochemical studies determined that AnkX directly mediates the covalent attachment of a phosphocholine moiety to Rab1. This phosphocholine transferase activity used CDP-choline as a substrate and required a conserved histidine residue located in the FIC domain of the AnkX protein. During infection, AnkX modified both Rab1 and Rab35, which explains how this protein modulates membrane transport through both the endocytic and exocytic pathways of the host cell. Thus, phosphocholination of Rab GTPases represents a mechanism by which bacterial FIC-domain-containing proteins can alter host-cell functions.

Legionella pneumophila is an intracellular pathogen that translocates proteins called effectors into the host-cell cytosol using a type IV secretion system called Dot/Icm⁸. The *Legionella* protein DrrA (also known as SidM) is an effector that targets the host GTPase Rab1 (refs 1–3, 5). Initially identified as a Rab1-specific guanine nucleotide exchange factor (GEF), recent studies showed that the amino-terminal region of DrrA has structural similarity to glutamine synthetase adenylyl transferase (GS-ATase) and shares the catalytically important sequence motif G-X₁₁-D-X-D, which enables DrrA to AMPylate the Tyr 77 residue in the class II switch region of Rab1B⁷. To determine if the *in vitro* activity described for DrrA is biologically relevant we examined whether the endogenous DrrA protein mediates Rab1 AMPylation when delivered into host cells during *Legionella* infection.

Cells were infected with a strain of *Legionella* that has a functional Dot/Icm system that delivers effectors into host cells (wild type) or an isogenic Δ dotA mutant that has a non-functional Dot/Icm system, and Rab1 protein was analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Fig. 1a). Two different modifications in the switch II region in Rab1 were detected after infection with wild-type *Legionella*. A fragment that corresponded to an AMPylated TITSSYYR peptide (mass to charge ratio, $m/z = 660.5$) was detected. Unexpectedly, a form of this peptide with an unknown moiety of 183 Da ($m/z = 578.5$) was also detected (Fig. 1a). The Δ dotA mutant revealed that both modifications required the delivery of effector proteins into host cells during infection. Thus, Rab1 is modified during *Legionella* infection by AMPylation and by a second unknown post-translational mechanism.

Cells were infected with mutant strains of *Legionella* deficient in effectors that could be involved in AMPylation of Rab1. In addition to

the Δ drrA mutant, a Δ ankX mutant of *Legionella* was examined. The AnkX protein contains a FIC domain, which for other bacterial effectors has been shown to have an enzymatic activity that promotes the AMPylation of small GTPases^{9–13}. When microinjected into mammalian cells the AnkX protein disrupts membrane transport in the secretory pathway and interferes with the sorting of transferrin from early endosomes, consistent with AnkX being an effector that disrupts the activities of host membrane transport proteins, potentially by Rab AMPylation¹⁴.

Rab1 AMPylation was not detected in the samples isolated from cells infected with the Δ drrA mutant, indicating that DrrA is the primary effector mediating Rab1 AMPylation *in vivo* (Fig. 1a). The unknown modification ($m/z = 578.5$) was detected after infection with the Δ drrA mutant, but was not detected after infection with the Δ ankX mutant. Thus, the unknown modification to Rab1 that occurs during infection requires AnkX. Defects in Rab1 modifications exhibited by these *Legionella* mutants were complemented upon the introduction of plasmids that restored DrrA and AnkX production (Supplementary Fig. 1a). MS/MS analysis revealed that the unknown 183 Da moiety was attached to Ser 79 of Rab1A, adjacent to the Tyr 80 residue AMPylated by DrrA (Fig. 1b). These residues correspond to Ser 76 and Tyr 77 in Rab1B.

Purified DrrA radiolabelled GST–Rab1 *in vitro* when ³²P- α -labelled ATP was used as a substrate, but no labelling was detected using ³²P- γ -labelled ATP, validating that DrrA mediates the attachment of AMP to Rab1 (Fig. 1c). The structurally distinct N-terminal region of DrrA was sufficient for AMPylation^{3,15–17}, and no AMPylation activity was detected for DrrA(340–533) or the DrrA(D110A,D112A) variant having the G-X₁₁-D-X-D adenylyl transferase domain inactivated. The effector AnkX was unable to efficiently AMPylate Rab1, the GTP-locked Rab1(Q70L) variant or the GDP-locked Rab1(S25N) variant, indicating that purified AnkX does not have robust Rab1 AMPylation activity (Fig. 1c).

The nature of the unknown modification to Rab1 requiring AnkX was investigated further. Cells were transfected with a plasmid encoding either AnkX or the variant AnkX(H229A), which has the essential histidine residue in the FIC domain changed to alanine. Roughly 70% of the Rab1 isolated from cells producing AnkX had the 183 Da moiety attached, whereas Rab1 isolated from cells producing the AnkX(H229A) protein was unmodified (Fig. 2a). Thus, AnkX is both necessary and sufficient to promote a novel post-translational modification to Rab1 by a process that requires a functional FIC domain.

For molecules <200 Da, the elemental composition can often be determined from a highly accurate mass measurement¹⁸. High-resolution MS measurements obtained for the modified Rab1 peptide isolated from cells producing AnkX revealed that the moiety attached to the Ser 79 residue had an accurate mass of 183.0661 Da (Supplementary Fig. 1b). This moiety did not match any known post-translational modifications, but when a metabolite database (<http://metlin.scripps.edu>) was searched a near perfect match was made to the molecule phosphocholine, which has an exact mass of 183.0660 Da. The protonated moiety attached to the Rab1 peptide was selected and further dissociated by multi-stage MS analysis (MS/MS/MS). The

¹Section of Microbial Pathogenesis, Yale University School of Medicine, Boyer Center for Molecular Medicine, Yale University, 295 Congress Avenue, New Haven, Connecticut, CT-06536, USA.

*These authors contributed equally to this work.

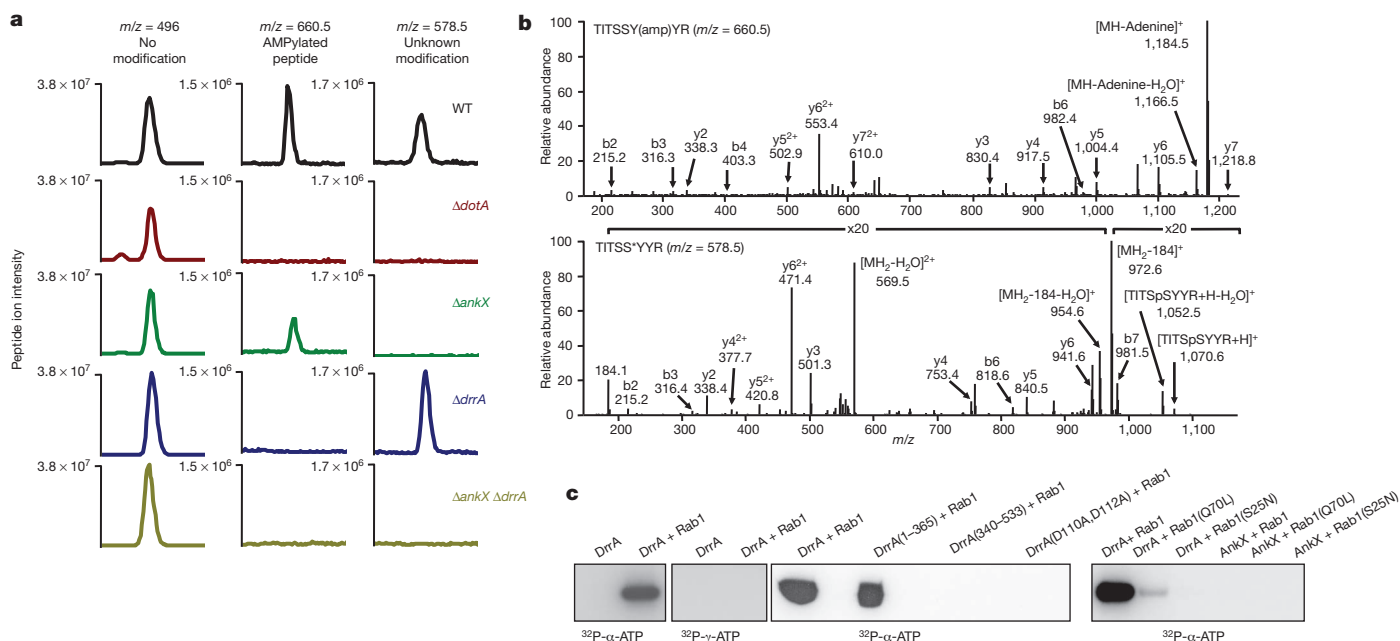


Figure 1 | *Legionella* infection mediates two different post-translational modifications to Rab1. **a**, 3×-Flag-Rab1 was isolated from HEK293 FcγRII cells after infection with the indicated *Legionella* strains. LC-MS/MS analysis produced the extracted ion chromatograms. The peak in each graph indicates the amount of Rab1 peptide TITSSYYR with no modifications ($m/z = 496.0$), peptide containing an AMP moiety ($m/z = 660.5$), and peptide with an unknown modification ($m/z = 578.5$). **b**, MS/MS spectra obtained for the AMPylated Rab1 peptide (TITSSY(amp)YR, top) and the Rab1 peptide with the unknown modification (TITSS*YYR, bottom) showing m/z values of their

fragments generated were of the sizes predicted for phosphocholine (Fig. 2b) and matched the MS/MS spectrum obtained following dissociation of a phosphocholine standard (Supplementary Fig. 2), suggesting that Rab1 is phosphocholinated by AnkX.

If AnkX were functioning directly as a phosphocholine transferase, the host molecule most likely to be used as a substrate in this reaction

would be CDP-choline, which is an intermediate used to synthesize phosphatidylcholine¹⁹. Indeed, phosphocholination of Ser 79 on Rab1 was detected for *in vitro* reactions containing CDP-choline and AnkX, but not in reactions containing DrrA (Fig. 2c) or the AnkX(H229A) protein (Supplementary Fig. 3). Increasing the amount of AnkX in the *in vitro* reaction resulted in higher levels of phosphocholinated Rab1

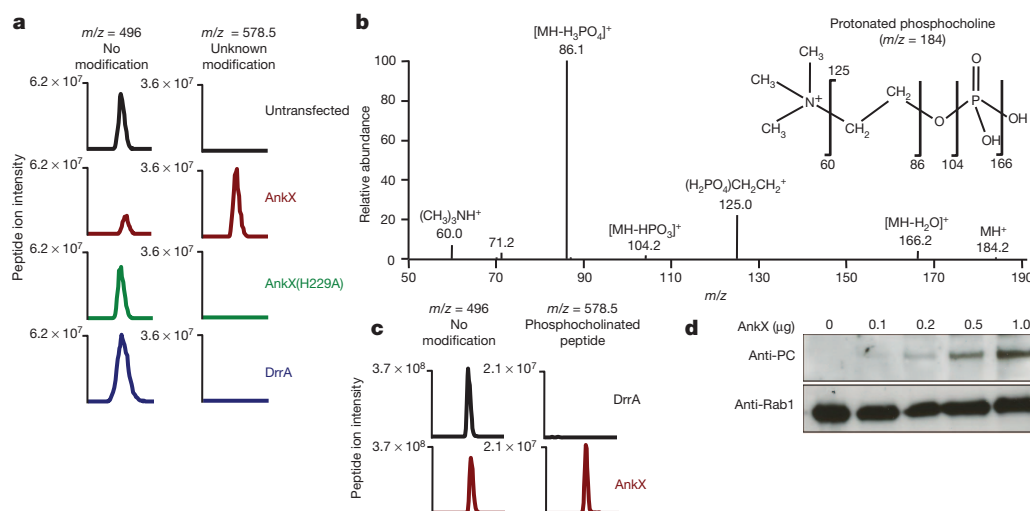


Figure 2 | The *Legionella* effector AnkX functions as a Rab phosphocholine transferase. **a**, LC-MS/MS analysis of 3×-Flag-Rab1 isolated from HEK293 cells that were either untransfected or transfected with a plasmid encoding GFP-tagged AnkX, AnkX(H229A) or DrrA. Extracted ion chromatograms indicate the amount of Rab1 peptide TITSSYYR with no modification ($m/z = 496$) and peptide with the unknown modification ($m/z = 578.5$). **b**, MS/MS/MS analysis on the Rab1 peptide TITSSYYR with the unknown modification. The $m/z = 184$ peak corresponding to the protonated moiety attached to the Rab1 peptide was selected and subjected to further dissociation.

Indicated are the fragments identified in the MS/MS/MS spectrum that matched the fragments predicted upon dissociation of the protonated phosphocholine molecule. **c**, The peak in each graph indicates the amount of Rab1 peptide TITSSYYR with no modifications ($m/z = 496$) and phosphocholinated peptide ($m/z = 578.5$) after *in vitro* incubation of Rab1 with either DrrA or AnkX in the presence of CDP-choline. **d**, Immunoblots from *in vitro* reactions that contained Rab1 and the indicated amounts of AnkX. Blots were probed to detect phosphocholinated Rab1 (anti-PC) and total Rab1 (anti-Rab1) in each reaction.

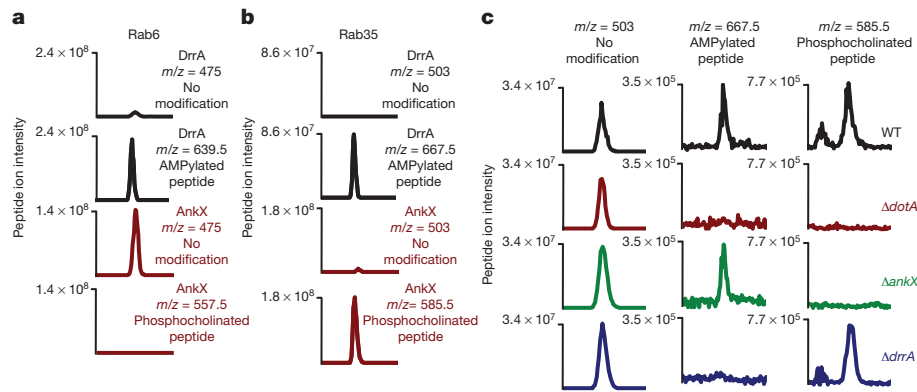


Figure 3 | AnkX and DrrA have overlapping but non-identical Rab specificities. **a**, Extracted ion chromatograms of 3×-Flag-Rab6 isolated from HEK293 cells producing either DrrA or AnkX. Shown are graphs for the unmodified Rab6 peptide SLIPSYIR ($m/z = 475$) and the AMPylated ($m/z = 639.5$) and phosphocholinated ($m/z = 557.5$) forms of this peptide. **b**, Extracted ion chromatograms of 3×-Flag-Rab35 isolated from HEK293 cells producing either DrrA or AnkX. Shown are graphs for the unmodified Rab35

peptide TITSTYYR ($m/z = 503.0$), the AMPylated peptide ($m/z = 667.5$) and the phosphocholinated peptide ($m/z = 585.5$). **c**, 3×-Flag-Rab35 was isolated from HEK293 FcγRII cells after infection with the indicated *Legionella* strains. The peak in each graph indicates the amount of Rab35 peptide TITSTYYR with no modifications ($m/z = 503$), peptide containing an AMP moiety ($m/z = 667.5$) and phosphocholinated peptide ($m/z = 585.5$). WT, wild type.

being detected by anti-phosphocholine immunoblot analysis (Fig. 2d), validating that Rab1 is phosphocholinated by AnkX.

Phosphocholinated proteins in the size range of Rab GTPases were detected in lysates from cells producing AnkX and were not observed in lysates from cells producing AnkX(H229A) (Supplementary Fig. 4a). Phosphatidylinositol 4-phosphate and phosphatidic acid levels were not affected in cells producing AnkX, suggesting that there is no indirect effect on phospholipid metabolism (Supplementary Fig. 4b). The intracellular pathogen *Coxiella burnetii* translocates a FIC-domain effector called CBU_2078 into host cells²⁰. Although proteins reacting with the anti-phosphocholine antibody were found in the size range of small GTPases from cell lysates producing CBU_2078, there was no evidence of Golgi fragmentation or endosome enlargement in these cells (Supplementary Fig. 4a, c). Thus, defects in host membrane transport in AnkX-producing cells probably results from phosphocholination of a specific subset of Rab GTPases, although it cannot be excluded that CDP-choline consumption might augment these effects.

The repertoire of Rab proteins that could be modified by *Legionella* effectors *in vivo* was investigated. Modifications to the Rab5 protein were not detected in cells producing either AnkX or DrrA (Supplementary Fig. 5). DrrA mediated the AMPylation of Rab6 on Tyr 82, but phosphocholination of Rab6 mediated by AnkX was not detected (Fig. 3a and Supplementary Fig. 5). Rab35 is a Rab1 family member that regulates the sorting of cargo from early endosomes, and interfering with Rab35 function results in enlarged early endosomes²¹. Importantly, specific perturbations in Rab35 function result in a cellular phenotype that closely mirrors the defects in endosome morphology observed in cells microinjected with purified AnkX¹⁴. Phosphocholinated Rab35 was detected in samples isolated from cells producing AnkX, and AMPylated Rab35 was detected from cells producing DrrA (Fig. 3b). During infection, phosphocholination of Rab35 required AnkX and AMPylation of Rab35 required DrrA (Fig. 3c). Thus, AnkX has specificity for Rab1 family members.

To test whether previously described cellular disruptions mediated by AnkX required the FIC-domain-dependent phosphocholine transferase activity (Fig. 4a), cellular phenotypes mediated by AnkX and the AnkX(H229A) mutant were compared. Disruption of the Golgi apparatus and a block in secretion of host alkaline phosphatase into the culture supernatant were observed in cells producing AnkX but not in cells producing AnkX(H229A) (Fig. 4b and Supplementary Fig. 7). Importantly, when the Ser 79 residue in Rab1 was changed to alanine, the variant protein was no longer phosphocholinated by AnkX (Supplementary Fig. 8a); however, Rab1(S79A) interfered with secretion of alkaline phosphatase when produced in cells (Supplementary Fig. 8b).

Thus, AnkX is targeting a residue in Rab1 that is critical for function. There was a significant increase in the number of cells containing enlarged early endosomes in cells producing AnkX compared to cells producing AnkX(H229A) (Fig. 4c and Supplementary Fig. 7), consistent with the function of Rab35 being perturbed by phosphocholination.

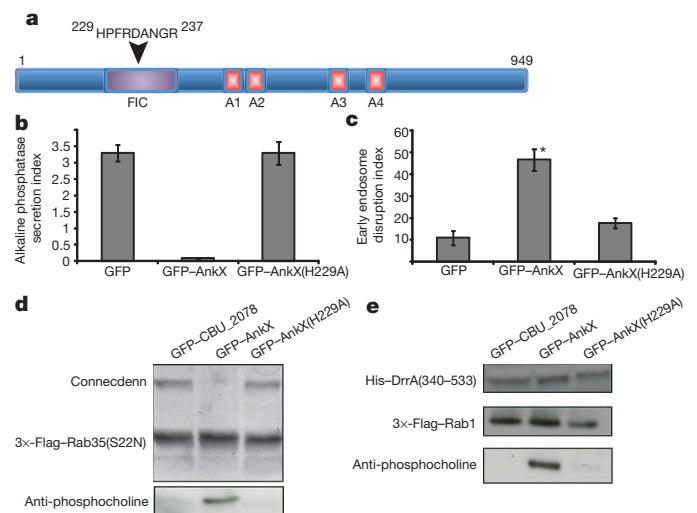


Figure 4 | AnkX-mediated phosphocholination modulates the function of Rab1 and Rab35. **a**, Schematic representation of the AnkX protein showing the location of the FIC domain and the four predicted ankyrin repeat homology domains (A1–A4). The amino acid sequence in a conserved region of the FIC domain containing the essential His 229 residue is shown. **b**, Secretion of alkaline phosphatase into the culture supernatant was measured for HEK293 cells producing either GFP, GFP–AnkX or GFP–AnkX(H229A) as indicated on the x-axis. Data are the mean \pm standard deviation (s.d.) calculated from three independent sample wells. **c**, The disruption of early endosomes was assessed in COS7 cells producing either GFP–AnkX, GFP–AnkX(H229A) or GFP alone after staining for EEA1 (mean \pm s.d., $n = 200$, $*P < 0.005$ compared to the GFP alone control). **d**, Binding of recombinant connectin to 3×-Flag–Rab35(S22N) isolated from cells producing either GFP–CBU_2078 or GFP–AnkX or GFP–AnkX(H229A) was assessed by co-precipitation. The Coomassie-stained SDS–PAGE gel indicates the amount of connectin and 3×-Flag–Rab35(S22N) in each precipitate. The anti-phosphocholine immunoblot indicates that Rab35(S22N) isolated from cells producing GFP–AnkX was phosphocholinated. **e**, Binding of purified His-tagged DrrA(340–533) to 3×-Flag–Rab1A isolated from cells producing either GFP–CBU_2078 or GFP–AnkX or GFP–AnkX(H229A) was assessed by co-precipitation. The immunoblots indicate the amounts of DrrA(340–533) and the levels of phosphocholinated 3×-Flag–Rab1A present in each precipitate.

Because GEF proteins are essential for Rab activation, the effect of phosphocholination on the binding of Rab-specific GEFs was analysed. The eukaryotic connecdenn proteins are the only known GEFs for Rab35 and are required for Rab35 function *in vivo*²¹. A pronounced defect in binding of connecdenn was observed for phosphocholinated Rab35 isolated from cells producing AnkX (Fig. 4d), which would explain why AnkX overproduction mimics the cellular phenotype observed when connecdenn has been silenced in mammalian cells²¹. By contrast, there was no defect in the binding of phosphocholinated Rab1 with the GEF domain of DrrA (Fig. 4e), similar to what has been observed for DrrA interactions with AMPylated Rab1 (ref. 7). Thus, post-translational modifications mediated by the effectors AnkX and DrrA modulate the function of Rab GTPases during infection by tailoring the repertoire of proteins that interact with the modified GTPase.

The characterization of DrrA and AnkX provides an example of *Legionella* having structurally distinct proteins with different biochemical activities that modulate the function of host vesicle transport proteins similarly. This concept of functional redundancy has been postulated but not shown clearly. The differences observed in the *in vivo* specificities shown by these two effectors, however, demonstrate that they are also likely to have roles in modulating Rab protein functions that do not overlap, which could explain why positive selection has led to the emergence of two different pathways to modify Rab protein function through post-translational modification.

The reaction mediated by AnkX has similarities to the AMPylation reaction demonstrated for other FIC domain proteins. Both reactions use a nucleotide-based substrate as the donor molecule that mediates the post-translational modification process (Supplementary Fig. 9). Interestingly, in the AMPylation reaction, hydrolysis of the phosphoanhydrous bond results in protein modification by the 5'-ribonucleotide of the donor substrate, whereas in the phosphocholination reaction the 5'-ribonucleotide is presumably released and the phosphocholine group is transferred to the polypeptide chain.

There are several examples of occasions where post-translational modifications introduced by bacterial toxins or effectors—which were thought to be the exclusive domain of pathogens—were discovered to represent mechanisms used to regulate eukaryotic cell functions. Curiously, the inclusion of a phosphocholine moiety in a protein structure has been indicated previously by studies examining peptides secreted by nematodes and from mammalian cells residing in the placenta^{22,23}. Thus, protein phosphocholination may also be used by eukaryotic organisms to modulate cellular functions.

METHODS SUMMARY

MS/MS analysis of Rab GTPases was conducted on immunoprecipitated proteins that were fractionated by SDS-PAGE, digested with trypsin in the gel, and extracted peptides were separated using nano-LC and electrosprayed directly onto a linear ion trap mass spectrometer (LTQ Velos, ThermoElectron) for MS and MS/MS analysis. All biochemical assays were conducted using purified proteins as described in Methods. The antibody TEPC-15 (Sigma) was used to detect phosphocholinated proteins by immunoblot analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 February; accepted 22 June 2011.

Published online 7 August 2011.

1. Murata, T. *et al.* The *Legionella pneumophila* effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nature Cell Biol.* **8**, 971–977 (2006).

2. Ingmundson, A., Delprato, A., Lambright, D. G. & Roy, C. R. *Legionella pneumophila* proteins that regulate Rab1 membrane cycling. *Nature* **450**, 365–369 (2007).
3. Schoebel, S., Oesterlin, L. K., Blankenfeldt, W., Goody, R. S. & Itzen, A. RabGDI displacement by DrrA from *Legionella* is a consequence of its guanine nucleotide exchange activity. *Mol. Cell* **36**, 1060–1072 (2009).
4. Machner, M. P. & Isberg, R. R. Targeting of host Rab GTPase function by the intravacuolar pathogen *Legionella pneumophila*. *Dev. Cell* **11**, 47–56 (2006).
5. Machner, M. P. & Isberg, R. R. A bifunctional bacterial protein links GDI displacement to Rab1 activation. *Science* **318**, 974–977 (2007).
6. Nagai, H., Kagan, J. C., Zhu, X., Kahn, R. A. & Roy, C. R. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**, 679–682 (2002).
7. Muller, M. P. *et al.* The *Legionella* effector protein DrrA AMPylates the membrane traffic regulator Rab1b. *Science* **329**, 946–949 (2010).
8. Ninio, S. & Roy, C. R. Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends Microbiol.* **15**, 372–380 (2007).
9. Yarbrough, M. L. & Orth, K. AMPylation is a new post-translational modification. *Nature Chem. Biol.* **5**, 378–379 (2009).
10. Yarbrough, M. L. *et al.* AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* **323**, 269–272 (2009).
11. Kinch, L. N., Yarbrough, M. L., Orth, K. & Grishin, N. V. Fido, a novel AMPylation domain common to Fic, Doc, and AvrB. *PLoS ONE* **4**, e5818 (2009).
12. Roy, C. R. & Mukherjee, S. Bacterial FIC proteins AMP up infection. *Sci. Signal.* **2**, pe14 (2009).
13. Worby, C. A. *et al.* The Fic domain: regulation of cell signaling by adenylation. *Mol. Cell* **34**, 93–103 (2009).
14. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**, 1651–1654 (2008).
15. Brombacher, E. *et al.* Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-phosphate-binding effector protein of *Legionella pneumophila*. *J. Biol. Chem.* **284**, 4846–4856 (2009).
16. Zhu, Y. *et al.* Structural mechanism of host Rab1 activation by the bifunctional *Legionella* type IV effector SidM/DrrA. *Proc. Natl Acad. Sci. USA* **107**, 4699–4704 (2010).
17. Suh, H. Y. *et al.* Structural insights into the dual nucleotide exchange and GDI displacement activity of SidM/DrrA. *EMBO J.* **29**, 496–504 (2009).
18. Gross, M. L. Accurate masses for structure confirmation. *J. Am. Soc. Mass Spectrom.* **5**, 57 (1994).
19. Li, Z. & Vance, D. E. Phosphatidylcholine and choline homeostasis. *J. Lipid Res.* **49**, 1187–1194 (2008).
20. Chen, C. *et al.* Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. *Proc. Natl Acad. Sci. USA* **107**, 21755–21760 (2010).
21. Allaire, P. D. *et al.* The connecdenn DENN domain: a GEF for Rab35 mediating cargo-specific exit from early endosomes. *Mol. Cell* **37**, 370–382 (2010).
22. Lovell, T. M. *et al.* Identification of a novel mammalian post-translational modification, phosphocholine, on placental secretory polypeptides. *J. Mol. Endocrinol.* **39**, 189–198 (2007).
23. Grabitzki, J., Ahrend, M., Schachter, H., Geyer, R. & Lochnit, G. The PCome of *Caenorhabditis elegans* as a prototypic model system for parasitic nematodes: identification of phosphorylcholine-substituted proteins. *Mol. Biochem. Parasitol.* **161**, 101–111 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Gulcicek and K. Stone at the Yale Keck Proteomic Facility for advice and providing the high-resolution MS/MS analysis, K. Reinish and X. Wu for providing purified connecdenn, L. Lucast and P. De Camilli for assistance with lipid analysis, X. Pan for assistance in constructing AnkX plasmids, A. Hubber for technical assistance and H. Newton for editorial assistance. This work was supported by an Anna Fuller Fellowship (S.M.), National Institutes of Health (NIH) Grants F32 AI082927 (J.M.), and NIH grants R01-AI064559, R01-AI048770 and Northeast Biodefense Center Grant U54-AI057158-Lipkin (C.R.R.).

Author Contributions S.M., X.L. and K.A. performed research. X.L. conducted the mass spectrometry analysis, K.A. generated HEK293 Fcγ 3 × Flag-Rab1 stable cell line, J.M. conducted studies on the CBU_2078 protein, and S.M. conducted all other research. S.M., X.L., J.M. J.E.G. and C.R.R. analysed results and wrote manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.R.R. (craig.roy@yale.edu).

METHODS

Cell culture and transfection. COS7 and HEK293 cells were grown in Dulbecco's modified Eagle medium (DMEM) from Gibco (Carlsbad) containing 10% heat-inactivated fetal bovine serum (FBS; Gibco). Cell lines were cultured at 37 °C in 5% CO₂. For transfection, COS7 or HEK293 cells were added to 12-mm coverslips in 24-well plates at a density of 2.5×10^4 cells per well. Cells were transfected with 0.5 µg of each plasmid. Cells were fixed 18 h after transfection in 4% PFA, permeabilized with 0.05% saponin and processed for immunofluorescence microscopy as described previously¹⁴.

Fluorescence microscopy. Digital images were acquired with a Nikon TE300 microscope using a $\times 100$ 1.4 N.A objective lens and a Hamamatsu ORCA-ER camera controlled by IP Lab software.

Alkaline phosphatase secretion assay. Secretion assays were performed as described²⁴. Briefly, HEK293 cells were co-transfected with a plasmid encoding secreted alkaline phosphatase and a plasmid encoding either GFP, GFP-AnkX or GFP-AnkX(H229A). After an 18 h incubation period, cells were washed with PBS and fresh medium was added. Cells were incubated for 10 h and then the alkaline phosphatase secretion index was determined by measuring the ratio of alkaline phosphatase protein secreted into the culture medium to the total amount of alkaline phosphatase protein in the assay well. The Tropix PhosphoLight System Kit (Applied Biosystems) was used to measure alkaline phosphatase activity and the Tecan Infinite M1000 plate reader with iControl Software was used to detect chemiluminescence. Data shown are the mean \pm s.d. from three independent samples for each condition. Results were validated in two independent assays.

Protein purification and *in vitro* AMPylation. His-tagged and GST-tagged proteins were purified as described previously¹. Purified GST-Rab1A (2 µg) was incubated with 0.4 µg of purified effector protein in buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl₂ and 0.1 mM GTPγS) and incubated for 1 h at 30 °C in the presence of 2 µCi of ³²P- α -labelled ATP or ³²P- γ -labelled ATP (Perkin Elmer). Labelled proteins were identified by autoradiography following SDS-PAGE. *In vitro* AMPylation results shown are representative of three independent assays.

Bacterial strains and plasmids. The *Legionella* strains were grown on charcoal yeast extract plates as described previously¹⁴. The parental strain (wild type) was *L. pneumophila* serogroup 1 strain Lp01, and the variant strains were all isogenic mutants described previously^{1,14,25,26}, with the exception of the *ΔankX ΔdrrA* double mutant, which was generated for this study using allelic exchange to introduce the *ΔankX* mutation into the *ΔdrrA* strain as described¹⁴. For all experiments, *Legionella* were isolated from charcoal yeast extract plates after growth for 2 days at 37 °C. The plasmid pEFGPC2 (Clontech) was used for all GFP fusion constructs, the plasmid pJB1806 was used to produce DrrA and AnkX in *Legionella*, pQE30 (Qiagen) was used for all His-tagged constructs, and pGEX2TK (GE LifeSciences) was used for all GST-tagged constructs. The Rab plasmids were constructed using cDNA encoding human RAB1A, RAB5A, RAB6A, RAB35 and canine RAB1A.

Cell lines and *Legionella* infection. HEK293 FCγRII cells²⁷ were used to create the HEK293 FCγRII 3×-Flag-Rab1A stable cell line. This cell line was used for assays examining Rab1 modifications during infection of host cells by *Legionella*. For each assay, cells grown to near confluency in two 10-cm dishes were infected with opsonized *Legionella pneumophila* strain LP01 (wild type) or the isogenic mutants at an estimated multiplicity of 100 bacteria to 1 host cell. After incubation for 0.5 h at 37 °C the cells were lysed in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl₂, 1% Triton X-100, 1 mM PMSF and protease inhibitor cocktail (Roche). Lysates were centrifuged at 17,900g and the post-nuclear supernatant was then incubated with Flag-antibody-coated beads (Sigma) at 4 °C for 1 h, washed and the Rab1 protein was eluted using Flag peptide. Samples were then processed for MS/MS analysis. This same approach was used to assay modifications to Rab35 after infection, except that the HEK293 FCγRII cell line was transfected with an expression plasmid producing 3×-Flag-Rab35 and then infected 18 h after transfection.

Golgi and early endosome disruption assay. The disruption of Golgi and early endosomes was assessed in COS7 cells producing GFP-AnkX, GFP-AnkX(H229A), GFP-CBU_2078 or GFP alone. Cells were fixed with 4% PFA 18 h after transfection and stained with mouse monoclonal antibodies specific for either GM130 or EEA1 (BD Transduction Laboratories) at a dilution of 1:200. Golgi and endosome morphology in GFP-positive cells was assessed by fluorescence microscopy. The Golgi was completely fragmented in nearly all the cells producing AnkX, whereas no significant fragmentation was observed compared to the background of untransfected cells in cells producing GFP-AnkX(H229A), GFP-CBU_2078 or GFP alone. The early endosome disruption index represents the percentage of cells producing the indicated GFP protein that showed an enlarged endosome phenotype as determined by EEA1 staining. The data represent the mean \pm s.d., from three independent replicates in which 200 cells were counted for each protein. *P* values were computed using Student's unpaired *t*-test. Data shown were validated in two independent experiments.

***In vitro* phosphocholination assay.** GST-tagged Rab1A (5 µg) attached to glutathione agarose was incubated with thrombin to remove the tag. Rab1A was incubated with 1 µg of purified effector protein in buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl₂, 1 mM ATP) and incubated for 1 h at 30 °C in the presence of 1 mM CDP-choline. Samples were boiled in SDS-loading buffer and the Rab1 protein was excised after SDS-PAGE and analysed by LC-MS/MS analysis. For the immunoblot analysis, the amounts of AnkX in each reaction varied from 0 µg to 1.0 µg, and the amount of modified Rab1 in each reaction was compared by immunoblot analysis using an antibody specific for phosphocholine (TEPC-15; Sigma) to detect modified Rab1 and an antibody specific for Rab1A (Santa Cruz Biotechnologies) to detect total Rab1. Data shown were validated in three independent experiments.

Detection of host proteins modified by AnkX and CBU_2048 *in vivo*. HEK293 cells in 24-well dishes were transfected with plasmids encoding GFP-CBU_2078, GFP-AnkX or GFP-AnkX(H229A) and cultured for 18 h. Cells were lysed in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl₂, 1% Triton X-100, 1 mM PMSF and protease inhibitor cocktail. The lysates were centrifuged at 17,900g and 50 µg of the supernatant was separated by SDS-PAGE for immunoblot analysis using the anti-phosphocholine-specific antibody TEPC-15 (Sigma).

GEF binding assays. To measure connectenn binding to modified Rab35, a plasmid encoding 3×-Flag-Rab35(S22N) was co-transfected into cells together with a plasmid encoding GFP-CBU_2078, GFP-AnkX or GFP-AnkX(H229A). The GDP-locked allele of Rab35 was used because it demonstrates enhanced connectenn binding. Cellular lysates were prepared as described earlier, and post-nuclear supernatant was incubated with 4 µg of purified connectenn (GenBank accession number NP_659414). 3×-Flag-Rab35(S22N) was precipitated using anti-Flag agarose (Sigma) and eluted from the beads using the 3×-Flag peptide (Sigma). The eluted proteins were resolved by SDS-PAGE gel and proteins were identified after the gel was stained with Coomassie brilliant blue dye. The locations of 3×-Flag-Rab35(S22N) and connectenn were determined by running purified connectenn and 3×-Flag-Rab35(S22N) in adjacent wells. To measure DrrA binding to modified Rab1, HEK293 cells that stably produce 3×-Flag-Rab1 were transfected with GFP-CBU_2078, GFP-AnkX or GFP-AnkX(H229A). 500 µg of post-nuclear supernatant was incubated with 0.5 µg of purified His-tagged DrrA(340–533). 3×-Flag-Rab1 was precipitated using anti-Flag agarose beads (Sigma). Immunoblot analysis was used to compare the amount of DrrA (anti-His), the amount of total 3×-Flag-Rab1 (anti-Flag, Sigma) and the amount of phosphocholinated Rab1 (TEPC-15; Sigma) after 2% of the total amount of 3×-Flag-Rab1 precipitated from each reaction was fractionated by SDS-PAGE. Data shown were validated in three independent experiments.

Mass spectrometry. 3×-Flag-Rab proteins were immunoprecipitated from cells using anti-Flag agarose beads, and separated by SDS-PAGE. The band corresponding to the 3×-Flag-Rab protein was excised from the gel and treated with DTT to reduce disulphide bonds and then alkylated with iodoacetamide (IAM). Trypsin digestion was allowed to occur overnight. Resulting peptides were extracted from gel matrix and then resuspended in aqueous buffer before final LC-MS/MS analysis. Nanoflow reverse-phase LC separation was carried out on a Proxeon EASY-nLC system (Thermo Scientific). The capillary column (75 µm \times 150 mm, PICOFRIT, New Objective) was packed in-house. A methanol slurry containing 5 µm, 100 Å Magic C18AQ silica-based particles (Microm BioResources) was forced to run through an empty capillary (with a frit in the end) using a pressurized device. The LC mobile phase was comprised of solvent A (97% H₂O, 3% acetonitrile (ACN) and 0.1% formic acid (FA)) and solvent B (100% ACN and 0.1% FA). The nano-LC separation was performed with the following gradient: B was increased from 7% to 35% in 40 min and then raised to 90% in 3 min and kept there for 10 min before going back to 100% A for column equilibration. At the moment when peptides were eluted from the capillary column, they were electrosprayed directly onto a linear ion trap mass spectrometer (LTQ Velos, ThermoElectron) for MS and MS/MS analysis. A data-dependent mode was enabled for peptide fragmentation with one full MS scan followed by collision induced dissociation (CID) of the ten most intense peptide ions. Dynamic exclusion was enabled to preclude repeated analyses of the same precursor ion. MS/MS scans were processed and searched using MASCOT (Matrix Science). The resulting peptide and protein assignments were filtered to keep only those identifications with scores above extensive homology. High-resolution MS/MS analysis was performed and the data was acquired by a high-resolution mass spectrometer (Orbitrap) by the Keck Proteomic Facility at Yale University. All LC-MS/MS data was validated by at least two independent experiments.

24. Kagan, J. C., Stein, M. P., Pypaert, M. & Roy, C. R. *Legionella* subvert the functions of Rab1 and Sec22b to create a replicative organelle. *J. Exp. Med.* **199**, 1201–1211 (2004).

25. Zuckman, D. M., Hung, J. B. & Roy, C. R. Pore-forming activity is not sufficient for *Legionella pneumophila* phagosome trafficking and intracellular growth. *Mol. Microbiol.* **32**, 990–1001 (1999).

26. Berger, K. H., Merriam, J. J. & Isberg, R. R. Altered intracellular targeting properties associated with mutations in the *Legionella pneumophila dotA* gene. *Mol. Microbiol.* **14**, 809–822 (1994).
27. Arasaki, K. & Roy, C. R. *Legionella pneumophila* promotes functional interactions between plasma membrane syntaxins and Sec22b. *Traffic* **11**, 587–600 (2010).

The genome sequence of Atlantic cod reveals a unique immune system

Bastiaan Star¹, Alexander J. Nederbragt¹, Sissel Jentoft¹, Unni Grimholt¹, Martin Malmstrøm¹, Tone F. Gregers², Trine B. Rounge¹, Jonas Paulsen^{1,3}, Monica H. Solbakken¹, Animesh Sharma⁴, Ola F. Wetten^{5,6}, Anders Lanzén^{7,8}, Roger Winer⁹, James Knight⁹, Jan-Hinnerk Vogel¹⁰, Bronwen Aken¹⁰, Øivind Andersen¹¹, Karin Lagesen¹, Ave Tooming-Klunderud¹, Rolf B. Edvardsen¹², Kirubakaran G. Tina^{1,13}, Mari Espelund¹, Chirag Nepal^{4,8}, Christopher Previti⁸, Bård Ove Karlsen¹⁴, Truls Moum¹⁴, Morten Skage¹, Paul R. Berg¹, Tor Gjøn¹⁵, Heiner Kuhl¹⁶, Jim Thorsen¹⁷, Ketil Malde¹², Richard Reinhardt¹⁶, Lei Du⁹, Steinar D. Johansen^{14,18}, Steve Searle¹⁰, Sigbjørn Lien¹³, Frank Nilsen¹⁹, Inge Jonassen^{4,8}, Stig W. Omholt^{1,13}, Nils Chr. Stenseth¹ & Kjetill S. Jakobsen¹

Atlantic cod (*Gadus morhua*) is a large, cold-adapted teleost that sustains long-standing commercial fisheries and incipient aquaculture^{1,2}. Here we present the genome sequence of Atlantic cod, showing evidence for complex thermal adaptations in its haemoglobin gene cluster and an unusual immune architecture compared to other sequenced vertebrates. The genome assembly was obtained exclusively by 454 sequencing of shotgun and paired-end libraries, and automated annotation identified 22,154 genes. The major histocompatibility complex (MHC) II is a conserved feature of the adaptive immune system of jawed vertebrates^{3,4}, but we show that Atlantic cod has lost the genes for MHC II, CD4 and invariant chain (Ii) that are essential for the function of this pathway. Nevertheless, Atlantic cod is not exceptionally susceptible to disease under natural conditions⁵. We find a highly expanded number of MHC I genes and a unique composition of its Toll-like receptor (TLR) families. This indicates how the Atlantic cod immune system has evolved compensatory mechanisms in both adaptive and innate immunity in the absence of MHC II. These observations affect fundamental assumptions about the evolution of the adaptive immune system and its components in vertebrates.

We sequenced the genome of a heterozygous male Atlantic cod (NEAC_001, Supplementary Notes 1 and 2), applying a whole-genome shotgun approach to 40× coverage (estimated genome size of 830 megabases (Mb), Supplementary Note 4 and Supplementary Fig. 2) using 454 technology (Supplementary Note 3). Two programs (Newbler⁶ and Celera⁷, Supplementary Notes 5 and 6) produced assemblies with short contigs, yet with scaffolds of comparable size to those of Sanger-sequenced teleost genomes (Supplementary Note 10 and Supplementary Fig. 8). Although fragmentation due to short tandem repeats is difficult to address (Supplementary Note 7), we resolved numerous gaps attributable to heterozygosity (Supplementary Note 8). The assemblies differ in scaffold and contig length (Table 1), although their scaffolds align to a large extent (Supplementary Note 9 and Supplementary Fig. 7). We obtained about one million single nucleotide polymorphisms (SNPs) by mapping 454 and Illumina reads from the sequenced individual to the Newbler assembly (Supplementary Note 11). Both assemblies cover more than 98% of the reads from an extensive transcriptome data set, indicating that the proteome is well represented (Supplementary Note 13). The assemblies are consistent with four

independently assembled bacterial artificial chromosome (BAC) insert clones (Supplementary Note 14 and Supplementary Fig. 9), and with the expected insert size of paired BAC-end reads (Supplementary Note 15 and Supplementary Fig. 10).

A standard annotation approach based on protein evidence was complemented by a whole-genome alignment of the Atlantic cod with the stickleback (*Gasterosteus aculeatus*), after repeat-masking 25.4% of the Newbler assembly (Supplementary Note 16 and Supplementary Table 6). In this way, 17,920 out of 20,787 protein-coding stickleback genes were mapped onto reorganized scaffolds (Supplementary Note 17). Additional protein-coding genes, pseudogenes and non-coding RNAs were annotated using the standard Ensembl pipeline. These approaches resulted in a final gene set of 22,154 genes (Supplementary Table 7). Comparative analysis of gene ontology classes indicates that the major functional pathways are represented in the annotated gene set (Supplementary Note 18 and Supplementary Fig. 11). We anchored 332 Mb of the Newbler assembly to 23 linkage groups of an existing Atlantic cod linkage map using 924 SNPs⁸ (Supplementary Note 19 and Supplementary Table 8). These linkage groups have distinct orthology to chromosomes of other teleosts, on the basis of the number of co-occurring genes, showing that the whole-genome shotgun assembly reflects the expected chromosomal ancestry (Fig. 1, Supplementary Note 20 and Supplementary Table 9).

Table 1 | Assembly statistics

	Number	Bases (Mb)	N50L (bp)*	N50 (n)†	ML (bp)‡
Newbler					
Contigs§	284,239	536	2,778	50,237	76,504
Scaffolds	6,467	611	687,709	218	4,999,318
Entire assembly	157,887	753	459,495	344	4,999,318
Celera					
Contigs§	135,024	555	7,128	19,938	117,463
Scaffolds	3,832	608	488,312	373	2,810,583
Entire assembly	17,039	629	469,840	395	2,810,583

*Minimum sequence length in which half of the assembled bases occur.

†Number of sequences with lengths of N50L or longer.

‡Maximum length.

§Contigs longer than 500 bp.

|| Scaffolds and unplaced contigs.

¹Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066, Blindern, N-0316 Oslo, Norway. ²Department of Molecular Biosciences, Centre for Immune Regulation, University of Oslo, Blindern, N-0316 Oslo, Norway. ³Bioinformatics Core Facility, Institute for Medical Informatics, Oslo University Hospital, Montebello, N-0310 Oslo, Norway.

⁴Department of Informatics, University of Bergen, N-5020 Bergen, Norway. ⁵Department of Natural Sciences and Technology, Hedmark University College, P.O. Box 4010, Bedriftsenteret, N-2306 Hamar, Norway. ⁶Department of Animal and Aquacultural Sciences, University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway. ⁷Department of Biology, Centre for Geobiology, University of Bergen, N-5020 Bergen, Norway. ⁸Computational Biology Unit, Uni Research AS, N-5020 Bergen, Norway. ⁹454 Life Sciences, 15 Commercial Street, Branford, Connecticut 06405, USA. ¹⁰Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ¹¹Nofima Marine, P.O. Box 5010, N-1430 Ås, Norway. ¹²Institute of Marine Research, P.O. Box 1870, Nordnes, N-5817 Bergen, Norway. ¹³Department of Animal and Aquacultural Sciences, CIGENE, Centre for Integrative Genetics, Norwegian University of Life Sciences, PO Box 5003, 1432 Ås, Norway. ¹⁴Faculty of Biosciences and Aquaculture, University of Nordland, N-8049 Bodo, Norway. ¹⁵Department of Pharmaceutical Biosciences, School of Pharmacy, University of Oslo, P.O. Box 1068, Blindern, N-0316 Oslo, Norway. ¹⁶Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin-Dahlem, Germany. ¹⁷Institute for Basic Sciences and Aquatic Medicine, School of Veterinary Sciences, N-0033 Oslo, Norway. ¹⁸Department of Medical Biology, Faculty of Health Sciences, University of Tromsø, N-9037 Tromsø, Norway. ¹⁹Department of Biology, PO Box 7803, University of Bergen, N-5020 Bergen, Norway.

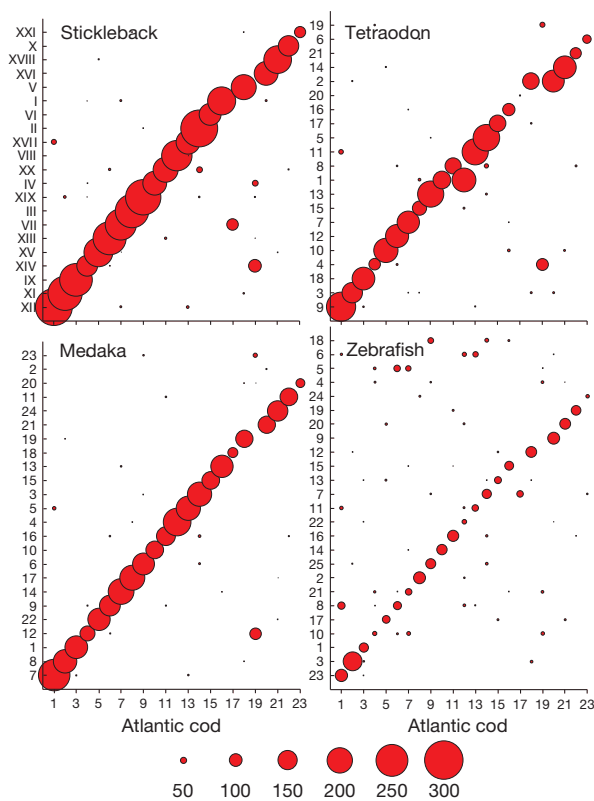


Figure 1 | Synteny between Atlantic cod and selected teleosts. The co-occurrence of orthologous genes (with a minimum of 50% sequence identity over 50% of the alignment, sphere size indicates the numbers of syntenic genes) in 23 Atlantic cod linkage groups⁸ (x-axis) reveals synteny with the chromosomes of four teleosts (y-axis). Several genes located on the stickleback chromosome XIV, tetraodon chromosome 4 and medaka chromosome 12 indicate a lineage-specific chromosomal rearrangement in Atlantic cod.

Well-studied haemoglobin polymorphisms in Atlantic cod are indicative of functional molecular adaptation to thermal variation^{9–12}. The genome contains nine α - and β -globin genes that are organized in two unlinked clusters, $\beta 5$ – $\alpha 1$ – $\beta 1$ – $\alpha 4$ and $\beta 3$ – $\beta 4$ – $\alpha 2$ – $\alpha 3$ – $\beta 2$ (refs 13, 14). We discovered an indel polymorphism of 73 base pairs (bp) in the intergenic promoter region of the $\alpha 1$ – $\beta 1$ globin pair (Fig. 2a and Supplementary Note 21). This promoter polymorphism occurs in highly significant linkage disequilibrium with two known polymorphic sites in the $\beta 1$ gene, the Val55Met and Ala62Lys substitutions¹, in eight Atlantic cod populations (Supplementary Note 22 and Supplementary Fig. 12). In fact, in the three most northern Atlantic populations and in both Baltic populations, the cod $\beta 1$ -globin gene predominantly occurs as a single homozygous genotype consisting of the long promoter and the Val 55–Ala 62 allele (Supplementary Table 10). By placing the two promoter variants in front of a luciferase reporter gene and transfecting the constructs into salmon kidney cells (Supplementary Note 23), we found that temperature and promoter type have a significant interaction effect (generalized linear model, $F_{2,36} = 7.85$, $P = 0.007$, Fig. 2b) and that the long promoter has twofold higher transcriptional activity compared to the short promoter at 15 °C and 20 °C. Increased globin synthesis of the Val 55–Ala 62 allele would compensate for its lower oxygen affinity^{10,11} at high temperatures. Thus, the promoter polymorphism provides a molecular compensatory mechanism that helps to maintain the total oxygen-carrying capacity¹⁵. The tight linkage between the two types of polymorphism provides a compelling example of the coevolution of structural and regulatory adaptation, and highlights the relationship between temperature and functional molecular variation in the haemoglobin system¹⁶.

The Atlantic cod immune system has unusual properties that set it apart from that of other teleosts: high levels of IgM¹⁷, a minimal antibody

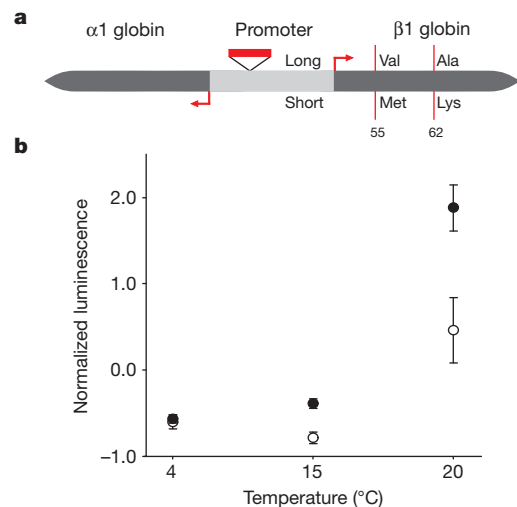


Figure 2 | Functional haemoglobin polymorphisms in Atlantic cod. a, Schematic of the head-to-head organized $\alpha 1$ and $\beta 1$ globin genes, the intergenic promoter region and transcription start sites (red arrows). A promoter polymorphism consisting of a 73-bp indel (red box) segregates in linkage disequilibrium with two amino-acid-substitution polymorphisms (vertical lines) at positions 55 and 62 in $\beta 1$ globin that affect its oxygen-binding affinity. This linkage disequilibrium results in two predominant haplotypes, long-Val-Ala and short-Met-Lys. b, Normalized luciferase luminescence ratios in salmon kidney cells. Cells were transfected using the long promoter (black circles) or the short promoter (white circles) and incubated at 4 °C, 15 °C or 20 °C ($n = 3$ for each treatment level). Error bars show 95% confidence intervals.

response after pathogen exposure^{5,17,18} and abundant phagocytic neutrophils in the peripheral blood^{19,20}. Despite speculation, the exact causes for these differences remain unknown⁵. We found that most genes involved in the vertebrate immune response are present in Atlantic cod (Supplementary Note 24, Supplementary Fig. 13 and Supplementary Table 11). Nevertheless, we did not find genes for the MHC II isoforms, their assembly and trafficking chaperone Ii²¹ and the MHC II-interacting protein CD4, which is essential for helper T-cell activation. By comparing a comprehensive set of vertebrate MHC II, CD4 and Ii sequences to the genome assemblies and all unassembled 454 and Illumina sequencing reads (a data set of about 49.5 gigabases), we detected a truncated pseudogene for CD4 (Supplementary Note 25), which is located in a region of conserved synteny (Supplementary Note 27 and Supplementary Fig. 18). No traces of MHC II and Ii were found in syntenous regions (Supplementary Note 27 and Supplementary Figs 16, 17, 19 and 20) and quantitative PCR (qPCR) targeting a conserved domain in MHC II did not amplify the target sequence (Supplementary Note 26 and Supplementary Fig. 15). The absence of MHC II and Ii, and the pseudogenic nature of CD4, show that Atlantic cod has lost the function of the classical pathway for adaptive immunity against bacterial and parasitic infections. Nevertheless, Atlantic cod deals adequately with its prevailing pathogen load in its natural ecological settings⁵. Previous transcriptional (complementary DNA) studies in Atlantic cod have indicated an expansion of the number of MHC I loci^{22,23}. By targeting the conserved MHC I $\alpha 3$ domain in genomic DNA using qPCR, we quantified more accurately the number of loci belonging to the teleost U-lineage²⁴ (Supplementary Note 28). Notably, Atlantic cod has about 100 classical MHC I loci, which is a highly expanded number compared to other teleosts (Fig. 3a). A phylogenetic analysis of teleost MHC I sequences supports the existence of two clades in cod (Fig. 3b and Supplementary Note 29). Within each clade, the mutation patterns show statistically significant signs of positive selection that are indicative of subfunctionalization. These findings indicate that loss of MHC II functionality has coincided with a more versatile usage of the cytosolic pathway of MHC I. Two different MHC I antigen-presentation pathways—the classical pathway and the alternative cross-presentation

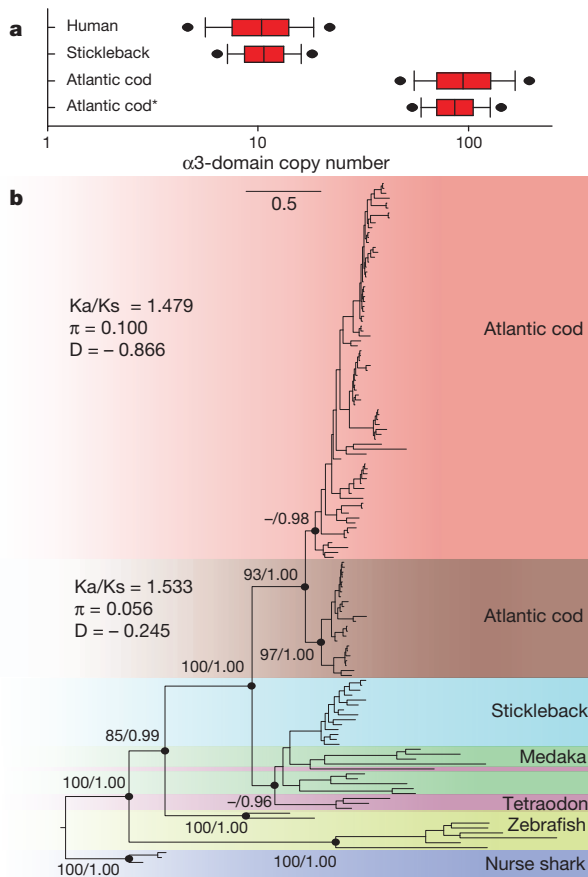


Figure 3 | MHC I diversity in Atlantic cod. **a**, Copy-number estimates of the MHC I $\alpha 3$ domain. Estimates are based on qPCR ratios (see Supplementary Note 28) of the MHC I $\alpha 3$ domain and a single-copy reference gene. For Atlantic cod, $\beta 2$ -microglobulin and topoisomerase III- α (*) were used as reference genes; for human and stickleback, $\beta 2$ -microglobulin was used. The estimates for human and stickleback agree with the expected number of $\alpha 3$ domains found in both reference genomes (Supplementary Table 15). Black dots indicate 95% confidence intervals calculated by bootstrapping ($n = 50,000$). **b**, Phylogeny of amino-acid sequences of MHC I $\alpha 1$ - $\alpha 3$ domains in teleosts. The Atlantic cod sequences are derived from cDNA and comprise classical U-lineage MHC I only. The other teleost sequences were obtained from Ensembl and NCBI, and contain classical and non-classical U-lineage MHC I. Alignments were visually inspected and corrected where necessary. Maximum likelihood (ML) values and Bayesian posterior probabilities (dots) support the main branches on the ML topology. Distance represents the number of substitutions per site (scale bar). The ratio of non-synonymous to synonymous variable sites (Ka/Ks), the average nucleotide diversity per site (π) and Tajima's D (D) were calculated for the two main clades in Atlantic cod.

pathway—can initiate immune responses in mammals²⁵. The cross-presentation pathway represents a structural and cellular modification of the MHC I machinery that allows activation of CD8⁺ T cells upon bacterial infection. The cytokine gene profile of Atlantic cod (Supplementary Table 11) supports the possibility of generating different subsets of CD8⁺ T cells that either provide direct protection or regulate other immune cells, and thus compensate for the loss of CD4⁺ T cells.

In addition to the MHC I expansion, we found an unusual composition of the highly conserved TLR families that have a fundamental role in the innate immune response and the initial detection of pathogens. Teleost TLR-encoding genes occur in well-supported phylogenetic clusters, most of which share functional properties with mammalian orthologues, although some are fish-specific²⁶. The Atlantic cod TLR genes form monophyletic groups within the known teleost functional groups (Fig. 4, Supplementary Note 30 and Supplementary Fig. 22). Genes for several TLRs that recognize bacterial surface antigens (TLR1, TLR2 and TLR5) are, however, absent, leaving only the teleost-specific

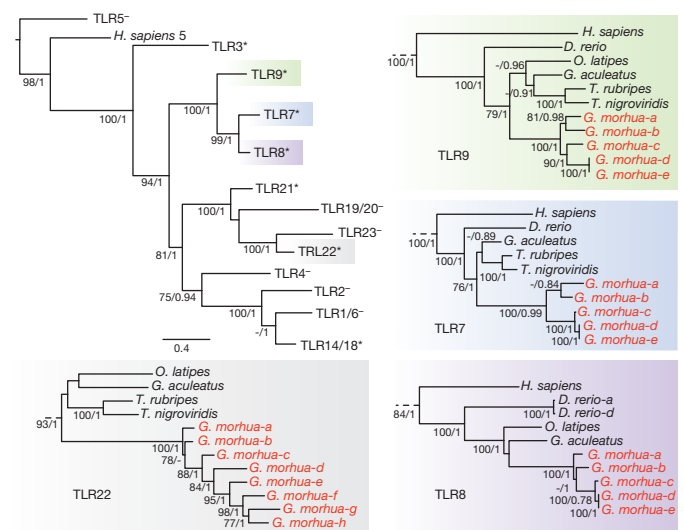


Figure 4 | Phylogeny of TLR families in Atlantic cod. TLR protein sequences were selected on the basis of the conserved Toll-IL-1 receptor (TIR) domain for Atlantic cod, including known sequences from stickleback, zebrafish, tetraodon, fugu, medaka and human as references. TLR clades with (*) or without (–) Atlantic cod sequences are denoted according to human or teleost orthologues (summary tree topology, top left panel). Distance represents the average number of substitutions per site (scale bar). ML values and Bayesian posterior probabilities greater than 75/0.75 support the ML topology. Detailed topologies of TLR7 (blue), TLR8 (purple), TLR9 (green) and TLR22 (grey) show gene expansions for Atlantic cod (red). Multiple TLR copies within species are subdivided by letters, and follow Ensembl nomenclature for *D. rerio*.

TLR14 and TLR18 as members of the TLR1 family in Atlantic cod. Moreover, several families of TLRs that recognize nucleic acids (TLR7, TLR8, TLR9 and TLR22) have markedly expanded, resulting in the highest number of TLRs found in a teleost so far. This TLR repertoire indicates that the Atlantic cod immune system relies relatively heavily on nucleic-acid-detecting TLRs to recognize bacterial pathogens. Notably, the gene expansion of TLR9 coincides with an expansion of interleukin-8 genes (IL-8, Supplementary Table 11). IL-8 is an important chemokine in the innate immune response and is directly induced by TLR9 in human neutrophils²⁷. The corresponding expansions of IL-8 and TLR9 indicate that this signalling cascade is particularly important in Atlantic cod.

The loss of MHC II function and lack of a CD4⁺ T-cell response represent a fundamental change in how the adaptive immune system is initiated and regulated in Atlantic cod. The marked expansion of MHC I genes and unusual TLR composition signify a shift of its immune system in handling microbial pathogens. An expanded MHC I repertoire in the presence of a non-polymorphic MHC II is found in an evolutionarily-distant vertebrate, the axolotl (*Ambystoma mexicanum*)^{28,29}. These observations indicate that anomalous immune systems (possibly analogous to that of Atlantic cod) have evolved independently. Additionally, we did not recover evidence for expressed MHC II, CD4 and Ii in the transcriptomes of three other gadoids, indicating that the unusual immune system is a derived characteristic of the gadoid lineage (Supplementary Tables 18 and 19).

We have provided the first annotated genome of a species that supports extensive fisheries and is on the verge of becoming an important aquaculture species. This work provides a major foundation for addressing key issues related to the management of natural Atlantic cod populations, such as the concept of fisheries-induced evolution, which dictates that selective harvesting can change the evolutionary trajectory of major life-history traits of natural populations³⁰. Moreover, our novel findings regarding the immune system will allow for more targeted vaccine development, aiding disease management and the process of domestication of Atlantic cod. These findings

change fundamental assumptions regarding the evolution of the vertebrate immune system.

METHODS SUMMARY

Detailed methods on the sequencing and assembly of data from genomic and transcriptomic origins; annotation, syntenic analyses, transfection experiments, bioinformatic analyses and phylogenetic analyses presented in this manuscript are described in the Supplementary Information.

Received 11 February; accepted 28 June 2011.

Published online 10 August 2011.

- Kurlansky, M. *Cod: A Biography of the Fish that Changed the World*. (Penguin, 1998).
- Johansen, S. D. *et al.* Large-scale sequence analyses of Atlantic cod. *New Biotechnol.* **25**, 263–271 (2009).
- Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Rev. Genet.* **11**, 47–59 (2010).
- Litman, G. W., Rast, J. P. & Fugmann, S. D. The origins of vertebrate adaptive immunity. *Nature Rev. Immunol.* **10**, 543–553 (2010).
- Pilström, L., Warr, G. W. & Stromberg, S. Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. *Fish. Sci.* **71**, 961–971 (2005).
- Miller, J. R., Koren, S. & Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315–327 (2010).
- Miller, J. R. *et al.* Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**, 2818–2824 (2008).
- Hubert, S., Higgins, B., Borza, T. & Bowman, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics* **11**, 191 (2010).
- Sick, K. Haemoglobin polymorphism in fishes. *Nature* **192**, 894–896 (1961).
- Brix, O., Thorkildsen, S. & Colosimo, A. Temperature acclimation modulates the oxygen binding properties of the Atlantic cod (*Gadus morhua* L.) genotypes *Hbl*1/1*, *Hbl*1/2*, and *Hbl*2/2* by changing the concentrations of their major hemoglobin components (results from growth studies at different temperatures). *Comp. Biochem. Physiol. A* **138**, 241–251 (2004).
- Andersen, O. *et al.* Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic cod populations. *Proc. R. Soc. B* **276**, 833–841 (2009).
- Petersen, M. F. & Steffensen, J. F. Preferred temperature of juvenile Atlantic cod *Gadus morhua* with different haemoglobin genotypes at normoxia and moderate hypoxia. *J. Exp. Biol.* **206**, 359–364 (2003).
- Borza, T., Stone, C., Gamperl, A. K. & Bowman, S. Atlantic cod (*Gadus morhua*) hemoglobin genes: multiplicity and polymorphism. *BMC Genet.* **10**, 51 (2009).
- Wetten, O. *et al.* Genomic organization and gene expression of the multiple globins in Atlantic cod: conservation of globin-flanking genes in chordates infers the origin of the vertebrate globin clusters. *BMC Evol. Biol.* **10**, 315 (2010).
- Gamperl, A. K., Busby, C. D., Hori, T. S. F., Afonso, L. O. B. & Hall, J. R. Hemoglobin genotype has minimal influence on the physiological response of juvenile Atlantic cod (*Gadus morhua*) to environmental challenges. *Physiol. Biochem. Zool.* **82**, 483–494 (2009).
- Weber, R. E. & Fago, A. Functional adaptation and its molecular basis in vertebrate hemoglobins, neuroglobins and cytoglobins. *Respir. Physiol. Neurobiol.* **144**, 141–159 (2004).
- Magnadottir, B., Gudmundsdottir, S., Gudmundsdottir, B. K. & Helgason, S. Natural antibodies of cod (*Gadus morhua* L.): specificity, activity and affinity. *Comp. Biochem. Physiol. B* **154**, 309–316 (2009).
- Solem, S. T. & Stenvik, J. Antibody repertoire development in teleosts—a review with emphasis on salmonids and *Gadus morhua* L. *Dev. Comp. Immunol.* **30**, 57–76 (2006).
- Rønneseth, A., Wergeland, H. I. & Pettersen, E. F. Neutrophils and B-cells in Atlantic cod (*Gadus morhua* L.). *Fish Shellfish Immunol.* **23**, 493–503 (2007).
- Øverland, H. S., Pettersen, E. F., Rønneseth, A. & Wergeland, H. I. Phagocytosis by B-cells and neutrophils in Atlantic salmon (*Salmo salar* L.) and Atlantic cod (*Gadus morhua* L.). *Fish Shellfish Immunol.* **28**, 193–204 (2010).
- Landsverk, O. J., Bakke, O. & Gregers, T. F. MHC II and the endocytic pathway: regulation by invariant chain. *Scand. J. Immunol.* **70**, 184–193 (2009).
- Persson, A. C., Stet, R. J. M. & Pilström, L. Characterization of MHC class I and β 2-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class I genes. *Immunogenetics* **50**, 49–59 (1999).
- Miller, K. M., Kaukinen, K. H. & Schulze, A. D. Expansion and contraction of major histocompatibility complex genes: a teleostean example. *Immunogenetics* **53**, 941–963 (2002).
- Dijkstra, J. M. *et al.* A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. *Immunogenetics* **59**, 305–321 (2007).
- Amigorena, S. & Savina, A. Intracellular mechanisms of antigen cross presentation in dendritic cells. *Curr. Opin. Immunol.* **22**, 109–117 (2010).
- Rebl, A., Goldammer, T. & Seyfert, H. M. Toll-like receptor signaling in bony fish. *Vet. Immunol. Immunopathol.* **134**, 139–150 (2010).
- József, L., Khreiss, T., El Kebir, D. & Filep, J. G. Activation of TLR-9 induces IL-8 secretion through peroxynitrite signaling in human neutrophils. *J. Immunol.* **176**, 1195–1202 (2006).
- Tournefier, A. *et al.* Structure of MHC class I and class II cDNAs and possible immunodeficiency linked to class II expression in the Mexican axolotl. *Immunol. Rev.* **166**, 259–277 (1998).
- Kaufman, J., Volk, H. & Wallny, H. J. A. “Minimal essential Mhc” and an “unrecognized Mhc”: two extremes in selection for polymorphism. *Immunol. Rev.* **143**, 63–88 (1995).
- Stenseth, N. C. & Dunlop, E. S. Evolution: unnatural selection. *Nature* **457**, 803–804 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the Research Council of Norway (FUGE program) to K.S.J. The authors wish to thank the following people and organizations: the 454 Life Science Sequencing Center (Branford, USA); the 454 and Illumina nodes of the Norwegian Sequencing Centre (University of Oslo); M. Egholm (formerly 454 Life Science); the Norwegian Metacenter for Computational Science (Notur) and the Norwegian Storage Infrastructure (Norstore); the Research Computing Services group, especially B. -H. Mevik, at the Center for Information Technology (University of Oslo); B. Walenz (Celera); the Canadian Cod Genomics and Broodstock Development Consortium; P. Olsvik, K. Lie and E. Holen at the Norwegian National Institute of Nutrition and Seafood Research (NIFES); J. Gaup and H. Bakke (CEES, University of Oslo); M. Kent (CIGENE, Norwegian University of Life Sciences); S. Bowman (Genome Atlantic); the FUGE bioinformatics platforms group, especially S. Grindhaug; I. Sandlie and O. B. Landsverk (Centre for Immune Regulation, University of Oslo); and Roche Norway.

Author Contributions DNA and RNA isolation, library construction and sequencing: A.T.-K., M.S., M.H.S., T.B.R., M.M., M.E., B.S., A.J.N. and J.T. Sanger BAC (end-) sequencing: H.K. and R.R. Assembly: A.J.N., B.S., A.S. and A.L. Linkage map analyses: K.G.T. and B.S. SNP analyses: K.G.T., P.R.B., S.L. and A.J.N. Annotation: J.-H.V., B.A. and S.S. Repeat analyses: B.S. Synteny analyses: J.P. and B.S. Haemoglobin analyses: Ø.A., O.F.W., B.S. and T.G. Bioinformatics: A.J.N., B.S., A.S., T.B.R., J.P., C.P., C.N., R.B.E., R.W., J.K., K.L., A.L., I.J., M.M., K.M., P.R.B., K.G.T. and M.H.S. Immune analyses: U.G., M.M., M.H.S., M.E., B.S., B.O.K., T.M., K.L., S.D.J. and T.B.R. Interpretation of immune results: U.G., T.F.G., S.J., B.S. and K.S.J. 454 contributions: L.D. Revisions: Ø.A., T.M., S.D.J., F.N., I.J., S.J., N.C.S. and S.W.O. Project initiation: S.W.O., I.J., F.N., S.L., N.C.S. and K.S.J. Project coordination: S.J. Consortium leader: K.S.J. This manuscript is dedicated to the memory of L. Pilström and R. J. M. Stet. Their research inspired our work to understand further the Atlantic cod immune system.

Author Information The unassembled sequencing reads and Newbler assembly have been deposited at ENA-EMBL under the accession numbers CAEA01000001–CAEA01554869. The annotation is available through Ensembl at <http://www.ensembl.org/index.html>. These and more resources are also available through <http://codgenome.no>. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share-Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.S.J. (k.s.jakobsen@bio.uio.no).